

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12Q 1/00, C07K 14/00, A61K 35/00,</b> <b>C12N 15/00</b>	<b>A2</b>	<b>(11) International Publication Number: WO 98/41648</b> <b>(43) International Publication Date: 24 September 1998 (24.09.98)</b>
<b>(21) International Application Number:</b> PCT/US98/05419 <b>(22) International Filing Date:</b> 19 March 1998 (19.03.98)  <b>(30) Priority Data:</b> 60/041,057 20 March 1997 (20.03.97) US  <b>(71) Applicant (for all designated States except US):</b> VARIAGEN-ICS, INC. [US/US]; One Kendall Square, Building 400, Cambridge, MA 02139-1562 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> HOUSMAN, David [US/US]; 64 Homer Street, Newton, MA 02159 (US). LEDLEY, Fred, D. [US/US]; 433 Grove Street, Needham, MA 02192 (US). STANTON, Vincent, P., Jr. [US/US]; 32 Royal Road, Belmont, MA 02178 (US).  <b>(74) Agents:</b> WARBURG, Richard, J. et al.; Lyon & Lyon LLP, Suite 4700, 633 West Fifth Street, Los Angeles, CA 90071-2066 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> TARGET GENES FOR ALLELE-SPECIFIC DRUGS  <b>(57) Abstract</b>  This disclosure concerns genetic targets which have been found to be useful for allele specific anti-tumor therapy. The strategy for such therapy involves the steps of: (1) identification of alternative alleles of genes coding for proteins essential for cell viability or cell growth and the loss of one of these alleles in cancer cells due to loss of heterozygosity (LOH) and (2) the development of inhibitors with high specificity for the single remaining alternative allele of the essential gene retained by the tumor cell after LOH. Particular categories of appropriate target genes are described, along with specific exemplary genes within those categories and methods of using such target genes.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	R	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DESCRIPTION

## TARGET GENES FOR ALLELE-SPECIFIC DRUGS

BACKGROUND OF THE INVENTION

This invention is concerned with the field of treatment of proliferative disorders, including malignant and nonmalignant diseases, and with transplantation.

Specifically, this invention is concerned with target genes for drugs that are useful for treating such diseases by providing allele-specific inhibition of essential cell  
5 functions.

The following information is provided to assist the understanding of the reader, none of that information is admitted to be prior art to the present invention.

The treatment of cancer is one of the most heavily investigated areas in biomedical research today. Although many anticancer drugs have been and continue to be  
10 discovered, there remains the immense problem of developing drugs that will be specifically toxic to cancer cells without killing normal cells and causing toxic, often permanent, damage to vital organs or even death. One common measure of the clinical usefulness of any anticancer drugs is its therapeutic index: the ratio of the median lethal dose (LD<sub>50</sub>) to the median effective dose (ED<sub>50</sub>) of the drug.  
15 With some cancer therapeutics this ratio is in the range of 4-6, or even 2-4, indicating a high risk of toxic side effects to the patient. Indeed, most anticancer drugs are associated with a high incidence of adverse drug events. The poor therapeutic index of most anticancer drugs not only limits the clinical efficacy of these drugs for the treatment of cancer, but limits their usefulness for treating  
20 many non-malignant, proliferative disorders.

A strategy for the development of anticancer agents having a high therapeutic

index is described in Housman, International Application PCT/US/94 08473 and Housman, INHIBITORS OF ALTERNATIVE ALLELES OF GENES ENCODING PROTEINS VITAL FOR CELL VIABILITY OR CELL GROWTH AS A BASIS FOR CANCER THERAPEUTIC AGENTS, U.S. Patent 5,702,890, issued December 30, 1997, which are hereby incorporated by reference in their entireties. As further described below, the method involves the identification of genes essential to cell growth or viability which are present in two or more allelic forms in normal somatic cells of a cancer patient and which undergo loss of heterozygosity in a cancer. Treatment of a cancer in an individual who is heterozygous with an allele specific inhibitor targeted to the single allele of an essential gene which is present in a cancer will inhibit the growth of the cancer cells. In contrast, the alternative allele present in non-cancerous cells (which have not undergone loss of heterozygosity) is able to express active product which supplies the essential gene function, so that the normal cells can survive and/or grow.

Cancer cells from an individual almost invariably undergo a loss of genetic material (DNA) when compared to normal cells. Frequently, this deletion of genetic material includes the loss of one of the two alleles of genes for which the normal somatic cells of the same individual are heterozygous, meaning that there are differences in the sequence of the gene on each of the parental chromosomes. The loss of one allele in the cancer cells is referred to as "loss of heterozygosity" (LOH). Recognizing that almost all, if not all, varieties of cancer undergo LOH, and that regions of DNA loss are often quite extensive, the genetic content of deleted regions in cancer cells was evaluated and it was found that genes essential for cell viability or cell growth are frequently deleted, reducing the cancer cell to only one copy. In this context, the term "deleted" refers to the loss of one of two copies of a chromosome or sub-chromosomal segment. Further investigation demonstrated that the loss of genetic material from cancer cells sometimes results

in the selective loss of one of two alleles of a certain essential gene at a particular locus or loci on a particular chromosome.

Based on this analysis, a therapeutic strategy for the treatment of cancer was developed, which will produce agents characterized by a high therapeutic index.

- 5 The strategy includes: (1) identification of genes that are essential (or conditionally essential) for cell survival or growth; (2) identification of common alternative alleles of these genes; (3) identification of the absence of one of these alleles in cancer cells due to LOH and (4) development of specific inhibitors of the single remaining allele of the essential gene retained by the cancer cell, but not the  
10 alternative allele.

#### SUMMARY OF THE INVENTION

- The utilization of inhibitors of alternative alleles, such as in the strategy described in Housman, *supra*, requires the provision of suitable target genes in order to identify such inhibitors and to implement corresponding diagnostic or therapeutic  
15 methods. Thus, as described below, the present invention identifies useful groups of genes which provide suitable target genes and further provides exemplary genes within those groups.

- Additionally, the present inventors determined that LOH occurs not only in cancers, but also in non-cancerous proliferative disorders, though the location and  
20 frequency of LOH differs in different diseases, and established a method by which such non-cancerous proliferative disorders can be treated. Noncancer proliferative disorders include, for example, atherosclerotic plaques, premalignant metaplastic or dysplastic lesions, benign tumors, endometriosis, and polycystic kidney disease. In each disease, the administration of such an inhibitor would have cytotoxic or  
25 antiproliferative effects on the abnormally proliferating cells that exhibited LOH and contained only the sensitive allele of the target gene, but would not be toxic to

normal cells that contain also the alternative allele.

In addition, it was found that specific inhibitors of alternative alleles of an essential gene would be useful in managing transplantation in instances where the alleles in a donor bone marrow differ from the alleles in the recipient. For  
5 example, administration of an inhibitor of an allele that was present in a donor bone marrow but not the recipient could be used to treat graft-versus-host disease, suppressing proliferation of the donor marrow without toxicity to the recipient. Alternatively, an inhibitor of an allele that is present in the recipient but not the donor bone marrow could be used to enhance engraftment by preferentially  
10 creating space in the recipient bone marrow for the graft without inhibiting proliferation of the engrafted donor marrow.

In this context, a "gene" is a sequence of DNA present in a cell that directs the expression of a "biologically active" molecule or "gene product", most commonly by transcription to produce RNA ("RNA transcript") and translation to produce  
15 protein ("protein product"). Both RNA and protein may undergo secondary modifications such as those induced by reacting with other constituents of the cell which are also recognized as gene products. The gene product is most commonly a RNA molecule or protein, or a RNA or protein that is subsequently modified by reacting with, or combining with, other constituents of the cell. Such  
20 modifications may result, for example, in the modification of proteins to form glycoproteins, lipoproteins, and phosphoproteins, or other modifications known in the art. RNA may be modified by complexing with proteins, polyadenylation, or splicing. The term "gene product" refers to any product directly resulting from transcription of a gene. In particular this includes partial, precursor, and mature  
25 transcription products (*i.e.*, RNA), and translation products with or without further processing, such as lipidation, phosphorylation, glycosylation, or combinations of such processing (*i.e.*, polypeptides).

The term "target gene" refers to a gene where the gene, its RNA transcript, or its protein product are specifically inhibited or potentially inhibited by a drug. In references herein to genes or alleles, the term "encoding" refers to the entire gene sequence, including both coding and non-coding sequences unless clearly indicated  
5 otherwise.

The term "allele" refers to one specific form of a gene within a cell or within a population, the specific form differing from other forms of the same gene in the sequence of at least one, and frequently more than one, variant sites within the sequence of the gene. The sequences at these variant sites that differ between  
10 different alleles are termed "variances", "polymorphisms", or "mutations". The term "alternative allele", "alternative form", or "allelic form" refers to an allele that can be distinguished from other alleles by having distinct variances at at least one, and frequently more than one, variant site within the gene sequence.

It is recognized in the art that variances occur in the human genome at  
15 approximately one in every 100-500 bases. At most variant sites there are only two alternative variances, wherein the variances involve the substitution of one base for another or the insertion/deletion of a short gene sequence. Within a gene there may be several variant sites. Alternative alleles can be distinguished by the presence of alternative variances at a single variant site, or a combination of  
20 several different variances at different sites. In this invention, inhibitors targeted to a specific allelic form or subset of the allelic forms of a gene can be targeted to a specific variance in a selected variant site, or to an allele comprised of a set of variances at different sites. In most but not all cases, the target specificity is based on a nucleotide or amino acid change at a single variance site.

25 The term "proliferative disorder" refers to various cancers and disorders characterized by abnormal growth of somatic cells leading to an abnormal mass of



tissue which exhibits abnormal proliferation, and consequently, the growth of which exceeds and is uncoordinated with that of the normal tissues. The abnormal mass of cells is referred to as a "tumor", where the term tumor can include both localized cell masses and dispersed cells. The term "cancer" refers to a neoplastic growth and is synonymous with the terms "malignancy", or "malignant tumor". The treatment of cancers and the identification of anticancer agents is the concern of particularly preferred embodiments of the aspects of the present invention. Other abnormal proliferative diseases include "nonmalignant tumors", and "dysplastic" conditions including, but not limited to, leiomyomas, endometriosis, benign prostate hypertrophy, atherosclerotic plaques, and dysplastic epithelium of lung, breast, cervix, or other tissues. Drugs used in treating cancer and other non-cancer proliferative disorders commonly aim to inhibit the proliferation of cells and are commonly referred to as antiproliferative agents.

"Loss of heterozygosity", "LOH", or "allele loss" refers to the loss of one of the alleles of a gene from a cell or cell lineage previously having two alleles of that gene. Normal cells contain two copies of each gene, one inherited from each parent. When these two genes differ in their gene sequence, the cell is said to be "heterozygous". The term heterozygous indicates that a cell contains two different allelic forms of a particular gene and thus indicates that the allelic forms differ at at least one sequence variance site. When one allele is lost in a cell, that cell and its progeny cells, comprising its cell lineage, become "hemizygous" for that gene or "partially hemizygous" for a set of genes, and heterozygosity is lost. LOH occurs in all cancers and is a common characteristic of non-malignant, proliferative disorders. In general, many different genes will be affected by loss of heterozygosity in a cell which undergoes loss of heterozygosity. In many cancers 10-40% of all of the genes in the human genome (there are estimated to be 60,000-100,000 different genes in the genome) will exhibit LOH. In the context of this invention, these terms refer preferably to loss of heterozygosity of a gene

which has a particular sequence variance in normal somatic cells of an individual such that there is loss of heterozygosity with respect to that particular sequence variance. Also preferably, these terms refer to loss of heterozygosity of a particular sequence variance that is recognized by an inhibitor that will inhibit one  
5 allele of the gene present in normal cells of the individual, but not an alternative allele.

Preferably, loss of heterozygosity occurs before clonal or oligoclonal expansion of cells associated with a condition or disease, for example, cancer or non-cancer proliferative disorder. Cancer is a "clonal" disorder, meaning that all of the cells  
10 in the cancer or tumor are the progeny, or lineage, of a single cell which undergoes malignant transformation. Since cancer is clonal, any loss of heterozygosity or allele loss that occurs during the process of malignant transformation will be uniformly present throughout the lineage of the initial transformed cell. This results in the cancer cells uniformly and consistently  
15 having only one allelic form of the gene which is present in two allelic forms in normal cells.

Some of the non-malignant proliferative conditions that exhibit LOH are "oligoclonal", meaning that unlike cancers and most benign tumors, there are multiple, independently arising clonal populations, with discrete LOH events in  
20 each of the individual clones. The alleles subject to LOH may vary from one clone to another. Therefore treatment of these conditions preferably utilizes inhibitors of at least two allelic forms. Thus, methods relating to such disorders can utilize alternative alleles of one gene and/or allelic forms of additional genes. Certain noncancer, proliferative disorders are considered to be precursors for  
25 cancer. Such disorders progressively exhibit LOH until a single cell within the lesion caused by abnormal proliferation undergoes transformation and clonal expansion to form a cancer. Because LOH occurs in the precancerous condition,

the present invention provides a method for preventing cancer by administering drugs that are selectively toxic to cells in which LOH involving a gene that is essential for cell survival or proliferation creates a genetic difference between cancer cells and normal cells. Since certain cancers are predictably associated with a high frequency of LOH in certain locations, for example segments of chromosomes 7,8,10,11,13,16, and 18 in prostate cancer, administration of an allele-specific drug that inhibits one allele that is within such a region, in a patient who is heterozygous for alternative forms of the gene, would kill cells that undergo LOH before cancer occurs. Preferably, in the context of this invention, LOH refers to loss of an allelic form of an essential gene in cells that are involved in cancer or noncancer proliferative disorders, which has sequence variants in a population of interest, in an individual whose normal somatic cells are heterozygous for sequence variants of that gene.

As pointed out above, an important aspect of methods for treating cancer or noncancer proliferative disorders utilizing LOH of essential genes is the identification of suitable essential genes for use as target genes. In accord with that requirement, this invention identifies certain useful groups or categories of essential genes, and provides, as examples, specific genes within those categories which are found to be suitable as targets for allele specific inhibitors, in particular for killing cancer cells or reducing the proliferation of cells in cancer or noncancer proliferative disorders. Thus, the present invention provides suitable target genes and methods of utilizing those genes in allele specific or variance specific targeting. Such targets are essential genes, which can include conditionally essential genes. As further described below, suitable target genes include those essential genes which encode gene products necessary for maintaining the level of a cellular constituent within the levels required for cell survival or proliferation, or which encode a gene product required for cell proliferation. If the level of activity of an essential gene product is reduced, the level of the corresponding cellular

constituent will not be properly maintained or the cell will be unable to perform the cellular functions required for cell proliferation. Confirmation that such a gene undergoes LOH in a neoplastic condition, *e.g.*, a cancer, and that there are at least two alleles of the gene in the population that differ in one or more variant  
5 positions, indicates that the gene is a useful potential target gene in this invention for the identification of allele specific inhibitors and in other aspects of the invention.

Certain useful groups of target genes are described in which the essential genes have been grouped according to the type of essential cellular function in which the  
10 gene products are involved. Thus, the gene product of each of the individual genes within each of the categories or subcategories is itself essential to the cell. In particular, the categories of genes, or cell functions shown in Table 1 (in the Detailed Description below) provide appropriate target genes. Particular exemplary target genes are also identified in Tables 1 and 2 and the Examples  
15 (including a GenBank accession number (or other sequence identifier as recognized by those skilled in the art) identifying the gene and providing a known sequence) which can be used for identifying allele specific inhibitors and for use in other aspects of this invention. Preferably the gene has the LOH frequency and at least one sequence variance in the gene has a heterozygosity rate in a population as  
20 indicated as preferable below, and occurs at only a single locus in the human genome.

An "essential" gene or gene product is one which is crucial to cell growth or viability. The terms "essential", "vital for cell viability or growth", or "essential for cell survival and proliferation" have the same meaning. A gene is essential if  
25 inhibition of the function of such a gene or gene product will kill the cell or inhibit its growth as determined by methods known in the art. Growth inhibition can be monitored as a reduction or preferably a cessation of cell proliferation.

Essentially can be demonstrated in a variety of different ways known in the art. Examples include, among others, generation of growth conditional mutants and identification of the affected genes, replacement of active genes with inactive mutants, cell fusion gene complementation analysis (see, *e.g.*, John Wasmuth, 5 "Chinese Hamster Cell Protein Synthesis Mutants", Ch. 14 in Molecular Cell Genetics, Michael Gottesman, ed. Wiley, New York, 1985), and insertion of genetic suppressor elements leading to growth arrest (Pestov & Lau, 1994, *Proc. Natl. Acad. Sci. USA* 91:12549-12533). Other ways include the identification of conditionally lethal mutants, *e.g.*, temperature sensitive mutants and determination 10 of the affected gene, genetic disruption of the gene by homologous recombination or other methods in organisms ranging from yeast to mice, inhibition of the gene by antisense oligonucleotides or ribozymes, and identification of the target of known cytotoxic drugs and other inhibitors. As further discussed below, the essentiality of a gene can depend on the conditions to which the cell is exposed.

15 Thus, unless otherwise indicated, the term "essential gene" includes both "generally essential genes" and "conditionally essential genes". "Generally essential genes" are those which are strictly essential for cell survival or growth, or which are essential under the conditions to which the cell is normally exposed. Typically such conditions are the normal *in vivo* conditions or *in vitro* conditions 20 which approximately replicate those *in vivo* conditions. Thus, in the methods described here utilizing essential genes, the method is carried out in conditions such that the gene product is required.

In connection with the determination of gene essentiality, it is generally recognized that the demonstration of essentiality of a gene in one organism is strongly 25 suggestive that the homologous gene will be essential in another organism. This is especially true for genes which have relatively high levels of sequence conservation across a broad range of organisms. Thus, the identification of essential genes in prokaryotes or in lower eukaryotes such as yeast is indicative of

the identification of corresponding homologous essential genes or gene classes in higher eukaryotes such as humans. Therefore, studies of essential genes for non-human organisms provides useful information on likely human essential genes; an example is the Stanford *Saccharomyces cerevisiae* Database: <http://genome-www.stanford.edu/cgr-bin/dbrun/SacchDB> which provides a catalog of essential genes in yeast. It should be recognized, however, that not all essential genes from lower organisms will have recognized homologues in humans. It should also be recognized that the essential genes for a particular organism will generally not be restricted to those for which homology can be shown to essential genes in other organisms. Thus, genes may be essential in humans that are not essential in lower organisms.

In addition to generally essential genes, it is also recognized in the art that environmental factors can cause certain genes to be essential that are not essential under other conditions (including usual culture conditions). For example, certain genes involved in intermediary metabolism are not essential if the cell or organism is supplemented with high concentrations of a particular nutrient or chemical entity, but if that nutrient or chemical entity is absent or present at low levels, the gene product is essential. In another example, the administration of a drug that inhibits one or more functions within the cell can cause other functions to be essential that are not essential in the absence of the drug. In another example, subjecting a cell to harsh physical agents, such as radiation, can cause certain genes to be essential that are not essential under normal conditions. Such genes are essential under certain conditions associated with the therapy of cancer. The demonstration that such genes are present in the population in more than one allelic form and are subjected to loss of heterozygosity in cancer or noncancer proliferative disorders makes such genes targets for allele specific drugs for the treatment of such disorders.

Thus, a gene is said to be “conditionally essential” if it is essential for cell survival or proliferation in a specific environmental condition caused by the presence or absence of specific environmental constituents, pharmaceutical agents, including small molecules or biologicals, or physical factors such as radiation.

- 5 The term “cellular constituent” refers to chemical entities that comprise the substance of a living cell. In preferred embodiments, the cellular constituent is a protein or modified protein. Also, in preferred embodiments, the cellular constituent is an inorganic ion, an organic compound such as a lipid, carbohydrate, amino acid, organic acid, nucleoside, DNA, or RNA, or modified
- 10 form of the preceding formed by the reaction of two constituents of the cell. In another embodiment, the constituent may comprise a structural element of the cell such as a membrane or cytoskeleton. In the preferred embodiment of this invention, cellular constituent refers to chemical entities, including compounds but also including simple ions, which are required for survival or proliferation of a
- 15 human cell.

Certain cellular constituents of a cell are synthesized by the cell while others are not synthesized by the cell but are taken into the cell from its environment. Within the cell, constituents engage in various reactions to form new constituents by intermediary metabolism, are modified to form new constituents, and are

20 preferentially compartmentalized in particular structures within the cell including, but not limited to, the nucleus, mitochondria, cytoplasm, or vesicles. Certain constituents are also specifically eliminated by the cell, or specific compartments within the cell, by degradation or excretion. In connection with cellular constituents, the term “maintaining the level” refers to maintaining the amount of

25 the chemical entity normally associated with a specific cellular compartment or compartments and involves the action of various cellular processes, including synthesis, production, compartmentalization, transport, modification, combining

of two or more constituents, polymerization, elimination, degradation, and excretion. It is recognized in the art that the failure to maintain the level of certain cellular constituents within normal levels results in cell death, for example, cell death may result from inappropriate levels of proteins, DNA, or RNA, 5 inappropriate levels of inorganic ions, inappropriate levels of organic compounds required for energy or other metabolic processes, or inappropriate intracellular structure. These examples are meant to be illustrative of the understanding of the meaning of the terms to those skilled in the art and not limiting.

10 In addition to the useful functional groups of essential genes described above, the present invention also provides useful groups of essential genes which are advantageous for allele specific targeting due to the genes undergoing LOH at certain frequencies in a disorder or other conditions and/or by having at least two allelic forms of the gene which appear in the population at particularly useful frequencies.

15 Thus, it is found that essential genes which undergo LOH in at least 10% of cases of a human cancer, and which exist in at least two allelic forms in a human population are advantageous targets. Preferably, the gene undergoes LOH in at least 20% of cases of a disorder, more preferably in at least 30%, still more preferably in at least 40%, and most preferably in at least 50% of such cases.

20 The LOH frequencies for a large number of different genetic markers for particular proliferative disorders are known in the art, and are used as indicators of the LOH frequency for neighboring essential genes. A number of LOH markers are provided in Fig. 3 (Loss of Heterozygosity Table). In one aspect of this invention, those essential genes which are located within about 20 megabases, 25 more preferably within about 10 megabases, and most preferably within about 5 megabases of an identified marker or tumor suppressor gene which undergoes



LOH in at least 10, 20, 30, 40, or 50% of cases of a proliferative disorder, are particularly useful as they will undergo LOH at similar frequencies as the marker gene.

The relative locations of a marker and an essential gene can also be described by genetic, rather than physical, map distances, therefore, in preferred embodiments, an essential gene of this invention is preferably within about 20 centimorgans, more preferably within about 15 centimorgans, still more preferably within about 10 centimorgans, and most preferably within about 5 centimorgans of such an LOH marker or tumor suppressor gene. In preferred embodiments, the target gene is located near a reported marker which undergoes LOH at a frequency of at least 10, 20, 30, 40, or 50% for a proliferative disorder. A number of such markers and the associated chromosomal locations are provided in Fig. 3. Even more preferably, essential genes which map to a locus bracketed by two such markers are appropriate potential target genes, as the essential gene very probably will also undergo LOH at similar high frequencies. Preferably both markers undergo LOH at frequencies of at least 10, 20, 30, 40, or 50% of cases of a cancer. Thus, confirmation that an essential gene, for example, a gene from one of the functional groups described above, or one of the particular exemplary genes, maps close to a marker as just described, indicates that the gene is an appropriate potential target. Identification of one or more sequence variances in that gene and/or in the corresponding gene products allows screening or design of such inhibitors for potential treatment.

A useful way to determine the frequency of loss of heterozygosity for a tumor cell based on the physical position of the gene on chromosomes within the human genome has been described by Vogelstein et al., 1989, *Science* 244:207-211. These authors describe a measure of allele loss termed Fractional Allele Loss (FAL) which quantifies the extent of LOH in cancer based on LOH determinations

over each informative chromosomal arm. FAL is determined by dividing the number of informative chromosomal arms which undergo LOH by the total number of informative chromosomal arms, *i.e.*, each chromosome/arm with at least one heterozygous locus in normal cells. Examples of such FAL

5 determinations are provided by Vogelstein et al., 1989 (FAL= 0.20 in colon cancer), and Cliby et al., 1993, *Cancer Research* 53:2393-2398 (FAL= 0.17 for low grade ovarian cancers, 0.40 for high grade ovarian cancers, 0.35 for all ovarian cancers).

These data indicate that genes on the chromosomal segment or

10 chromosomal arm that is commonly lost in a cancer or non-cancer proliferative disorder are potential target genes. In preferred embodiments, the target gene is located on a chromosomal arm which is reported in the art or shown herein to contain a locus or loci which undergoes LOH at a frequency of at least 15, preferably at least 20%, still more preferably at least 25%, and most preferably at

15 least 30, 40, or 50% in a proliferative disorder. As noted above, the frequency of LOH for a chromosomal arm is often utilized in calculating an average fraction of allele loss (FAL). Thus, a high LOH frequency for an arm or portion of an arm indicates that particular genes in the relevant chromosomal region will also undergo LOH at a comparable frequency, and thus define useful target genes.

20 Preferably the target genes are those which are located on particular chromosomal arms which commonly undergo tumor-related LOH. In particular, these human chromosomal arms include 1p, 1q, 3p, 5q, 6p, 6q, 7q, 8p, 9p, 9q, 10q, 11p, 11q, 13q, 16q, 17p, 17q, 18p, 18q, and 22q. It is recognized that the LOH frequency is not uniform for all positions along an arm of a particular chromosome, however

25 such LOH frequencies provide a strong indicator for LOH frequency at a potential target gene. Thus, mapping of an essential gene to these chromosomal arms or to high frequency LOH regions on these arms indicates that the gene is a potential target. Confirmation of the LOH of the particular gene and of the presence of at

least one sequence variance, and therefore of individuals heterozygous for such variances, indicates that the gene can be used for the identification of inhibitors targeting allelic forms of the gene which have a particular variance or variances and in the other aspects of this invention.

- 5 The term "high frequency LOH chromosomal region" refers to a chromosomal region which undergoes LOH at a frequency as indicated above, and include high frequency LOH chromosomal arms (at least 15% FAL), regions within the genetic or physical map distances indicated above of a chromosomal marker or tumor suppressor gene which undergoes LOH at a frequency as indicated above (at least  
10 10%).

In connection with the location of a potential target gene with respect to a marker or tumor suppressor gene, the term "proximity" means that the target gene is located within a genetic or physical map distance of the reference gene or marker as stated above.

- 15 The present invention is aimed, in part, at treating cancer or proliferative disorders of any type in which LOH of an essential gene occurs at a frequency as indicated above. For example, this includes but is not limited to cancers and noncancer proliferative disorders provided in Tables 2 and 3 and Figure 3, or otherwise described herein. Table 2 and Fig. 3 describe a number of cancers for  
20 which LOH at substantial frequencies has been described in the art. Therefore, identification of an essential gene which maps to the LOH regions for a particular proliferative disorder, as described by genetic or physical mapping or by residence on a chromosomal arm or smaller region of an arm which is shown to undergo LOH, at high frequency in a proliferative disorder, identifies a potential target  
25 gene. Identification of sequence variances in that gene, such that normal somatic cells of individuals in a population are heterozygous for a variance and thus

contain two different alleles, confirms that the gene is a potential target. The target gene, its RNA transcript or protein product can then be used as targets for allele-specific inhibitors for treating the proliferative disorder or other uses as described in the aspects of this invention.

- 5 A further indication of useful target genes is provided by tumor-specific LOH of essential genes associated with tumor suppressor genes. LOH in certain cancers or noncancer proliferative disorders is frequently associated with specific chromosomal arms. This association is believed to be due, in many cases, to the presence of tumor suppressor genes located on those particular chromosomal
- 10 arms, the loss of which eliminates the tumor suppressor function and contributes to the transformation of the cell. Consequently, essential genes which map near such a tumor suppressor gene are potential target genes for this invention. Preferably, the essential gene maps within a physical or genetic map distance as described above for LOH markers. As for the above categorization aspect, the
- 15 LOH for a particular gene preferably is at least 10, 20, 30, 40, or 50% for a tumor, such as the cancers and types of cancers identified in Tables 2 and 3 and in Fig. 3. It should be noted that tumor suppressor genes themselves are rarely essential for cell survival or proliferation and not likely to be preferred targets for this invention.
- 20 Another group of essential genes which are potentially useful as target genes are those which are present in the population in at least two alternative forms or alleles containing one or more sequence variations, where the alternate forms occur at frequencies such that at least 10% of a population is heterozygous (*i.e.*, have two alternative forms of the gene), preferably so that at least 20%, more
- 25 preferably at least 30%, and most preferably at least 40% are heterozygous. The term "heterozygote frequency" refers to the fraction of individuals in a population who have two alternative forms of a gene, or particular variances within a gene, in

their normal, somatic cells and are therefore heterozygous.

The term "allele frequency" refers to the fraction (or frequency of occurrence) of a specific allele as compared to all alleles in a population. It is recognized in the art that the heterozygote frequency and allele frequency are related and, for certain alleles, can be described by Hardy Weinberg equilibrium calculations. It will also be recognized that sequence variances that occur at high frequency in the population are commonly not deleterious to the health of the individuals who carry these genes and are commonly not disease genes or mutations that are associated with disease.

- 10 Methods for determining the heterozygote frequency or allele frequency or determining the number of individuals who are heterozygous for specific variances are known in the art, including but not limited to methods such as restriction fragment length polymorphism, hybridization of sequence specific nucleic acid probes to DNA or RNA sequences which include a sequence variance site, DNA  
15 sequencing, or mass spectrometry of amplified sequence fragments containing a sequence variance site. Methods that are useful for the discovery of genetic variances can also be used including, but not limited to, methods such as methods such as the SSCP technique (see Example 28), Enzymatic Mutation Detection technique (see Example 29), Denaturing Gradient Gel Electrophoresis, or  
20 sequencing. Identification of such genes which have sequence variances that are common in the general population and for which 10%, 20%, 30%, or 50% of the population are heterozygous for that gene provides genes which are particularly likely to be useful target genes for allele specific inhibition in this invention. Confirmation that the gene undergoes LOH at a useful frequency in a proliferative  
25 disorder, preferably in at least 10, 20, 30, 40, or 50% of cases of such a disorder indicates that the gene is useful as a potential target for identifying allele specific inhibitors for the treatment of proliferative disorders and in other aspects of this

invention.

Exemplary genes described herein are shown to contain numerous sequence variances which are present in human populations. While some sequence variances and alleles are common throughout diverse human populations, it is  
5 recognized in the art that the allele frequency of different genes will vary in different populations. For example, allele frequencies have been shown to differ between populations comprised of individuals of different races, populations comprised of individuals from different countries, populations comprised of individuals from different regions, populations comprised of individuals with  
10 common ethnic background, and even populations comprised of individuals from different religions. Alleles that are common in one population, may be rare in another. While the allele frequency of any particular gene may vary in different populations, the genes that are described below are those that occur such that at least 1% or 5% of a population is heterozygous for the sequence variance,  
15 preferably so that at least 10% or 20%, more preferably at least 30%, and most preferably at least 40% are heterozygous in a specific population that may be treated with inhibitors to treat cancer or other proliferative disorder in that population. Once a specific variance is identified in a certain gene, the allele frequency in any specific population can be easily determined using methods  
20 known in the art including the use of allele-specific hybridization probes, sequencing, or specific PCR reactions.

In this regard, "population" refers to a geographically, ethnically, or culturally defined group of individuals, or a group of individuals with a particular disease or a group of individuals that have proliferative diseases that may be treated by the  
25 present invention. Thus, in most cases a population will preferably encompass at least ten thousand, one hundred thousand, one million, ten million, or more individuals, with the larger numbers being more preferable. In special

circumstances, diseases will occur with high frequency in specific geographical regions or within specific familial, racial, or cultural groups, and a relevant population may usefully be considered to be a smaller group.

In the context of this invention, an alternative allele, or other reference to an  
5 appropriate target for the inhibitors of this invention refers to a form of a gene  
which differs in base sequence from at least one other allele or allelic form of the  
same gene. Usually, though not necessarily, the allelic forms of a gene will differ  
by, at most, several bases and may have only a single base difference (*i.e.*, a  
single sequence variance). The allelic forms, however, are ones which contain at  
10 least one sequence variance which appears in somatic cells of a population at an  
appreciable frequency, such that preferably at least 1%, more preferably at least  
5%, still more preferably at least 10%, and most preferably at least 20% of the  
population are heterozygous for that specific sequence variance. This  
advantageously allows the convenient identification of potential patients, because  
15 an appreciable fraction of the population, and therefore also of the cancer patients  
will be heterozygous for sequence variances of the specific gene. In the context of  
this invention, different alleles need not result in different observable phenotypes  
under normal conditions. Preferably, a particular sequence variance produces no  
phenotypic effect on the physical condition of an individual having that variance  
20 until the variance is targeted by an allele specific inhibitor.

In connection with allele specific inhibitors and the methods of this invention, the  
terms "allelic form" or "alternative form of the target gene" or "sequence  
variance within the target gene" refer to either or both of the gene or a product of  
that gene including the RNA transcript or protein product. Thus, a particular  
25 inhibitor may act in an allele specific manner (which will often be variance  
specific) at any of those levels and preferably the inhibitor is targeted to a  
particular sequence variance of the specific allelic form.

As indicated above, two different allelic forms of a gene will have at least a one nucleotide difference in the nucleotide sequence of the gene. The difference can be of a variety of different types, including base substitution, single nucleotide insertion or deletion, multiple nucleotide insertion or deletion, and combinations  
5 of such differences. Thus, two allelic forms are sequence variants and will have at least one sequence variance, which refers to the sequence difference, between the allelic forms. However, there may also be more than one sequence variance between two allelic forms. The location of a sequence variance in a gene sequence is a "sequence variance site." This description applies to both the DNA and RNA  
10 sequences, and similarly applies to a polypeptide sequence encoded by the gene, differences in the amino acid sequence of the polypeptide, and the location in the polypeptide chain of the sequence differences. As a particular gene may have more than one sequence variance site, more than two allelic forms may exist in a population, for example, see Fig. 1 for exemplary target summaries showing  
15 multiple sequence variance sites.

Sequence variances can involve a difference in the sequence in which any of the four bases: adenine, guanine, thymidine (uracil in the context of RNA), or cytosine are substituted with another of the four bases or a change in the length of the sequence. Different classes of variances are recognized in the art.  
20 "Deletions" are variances in which one or more bases are missing from the sequence. "Insertions" are variances in which one or more bases are inserted into the sequence. It will be evident that the terms deletion and insertion refer to the variance in one sequence relative to another. "Transitions" are variances that involve substitution of one purine for the other or one pyrimidine for the other.  
25 "Transversions" are variances that involve substitution of a purine for a pyrimidine or a pyrimidine for a purine. Certain sequence variances can interfere with the normal function of the gene or its gene product and can be associated with disease; such variances are commonly referred to as mutations. Most



variances present in human populations are not associated with disease and are “normal” variants of the gene; such variances are commonly referred to as polymorphisms. In the present invention, specific variances are described from each of the classes described above in genes that are essential for cell survival or proliferation that can be the targets for allele-specific inhibitors for the treatment of cancer or noncancer proliferative disorders.

This invention provides inhibitors which are specific for at least one, but not all, allelic forms of a gene that encodes a gene product essential to cell growth or cell viability, for genes belonging to the specified categories of genes. The inhibitor may be active on the gene or gene product including the RNA transcript, protein product, or modifications thereof. Exposure to the inhibitor inhibits proliferation or kills cells which have undergone LOH of genes that are not inhibited by the drug and contain only an allelic form of the essential gene, its RNA transcript, or its protein product against which the inhibitor is targeted. Normal cells which contain two alternative alleles of the target genes, one of which is not inhibited by the specific inhibitor, are spared from the toxic effects of the inhibitor because the remaining activity of the allele which is not inhibited by the inhibitor is adequate to permit continued cell viability and growth. This differential effect of the inhibitor on cells with LOH of a targeted gene (*e.g.*, a cancer cell) and normal cells accounts for the high therapeutic index of the inhibitors of this invention for the treatment of cancer or non-cancerous, proliferative disorders characterized by LOH. Toxicity of the inhibitor to normal cells is therefore low, compared to most currently available anticancer and antiproliferative agents.

Thus, in accord with the strategy and target genes indicated above and described in the Detailed Description of the Preferred Embodiments, in a first aspect the invention provides methods for identifying inhibitors potentially useful for treatment of a proliferative disorder, *e.g.*, cancer. Such inhibitors are active on

- specific allelic forms of target genes as identified herein. The method involves determining at least two allelic forms of such a gene encoding an essential gene product, and testing a potential allele specific inhibitor to determine whether the potential inhibitor is active on, *e.g.*, inhibits expression of, at least one of the
- 5 allelic forms, but not all of those forms. If the potential inhibitor inhibits only a subset of the allelic forms of the particular essential gene, then it is an allele specific inhibitor. Preferably the difference in activity of the inhibitor for different allelic forms is between allelic forms which have a sequence variance at a particular site.
- 10 In many, or even most, cases an allele specific inhibitor discriminates between two allelic forms due to a particular single sequence variance between the allelic forms of the target gene. For example, ribozymes which target a single sequence variance site will preferentially cleave only one of the sequence variants for a particular single nucleotide variance. In this case, sequence variances at other
- 15 sites will generally not affect the cleavage. In the Detailed Description of the Invention specific examples of proteins, small molecules, and oligonucleotides providing allele specific inhibition based on single sequence variances are described. Thus, in preferred embodiments an allele specific inhibitor discriminates between two allelic forms by discriminating a single sequence
- 20 variance. As previously indicated, inhibitors can be targeted to either the nucleic acid or a polypeptide (where a nucleotide change results in an amino acid change). In particular embodiments, the allele specific inhibitor will recognize more than one linked sequence variances within a specific allele.

- An "allele specific inhibitor" or "variance specific inhibitor" is a drug or inhibitor
- 25 that inhibits the activity of one alternative allele of a gene to a greater degree than at least one other alternative allele. The difference in activity is commonly determined by the dose or level of a drug required to achieve a quantitative degree

- of inhibition. A commonly used measure of activity is the IC<sub>50</sub> or concentration of the drug required to achieve a 50% reduction in the measured activity of the target gene. Preferably an allele specific inhibitor will have at least twice the activity on the target allelic form than on a non-target allelic form, more
- 5 preferably at least 5 times, still more preferably at least 10 times, and still more preferably at least 50 times, and most preferably at least 100 times. This can also be expressed as the sensitivities of the different allelic forms to the inhibitor. Thus, for example, it is equivalent to state that the target allelic form is most
- 10 preferably at least 100 times as sensitive to the inhibitor as a non-target allelic form. The activity of an inhibitor can be measured either *in vitro* or *in vivo*, in assay systems that reconstitute the *in vivo* system, or in systems incorporating selected elements of the complete biological system. For use in inhibiting cells containing only the target allelic form rather than cells containing at least one non-targeted allelic form, the difference in activity is preferably sufficient to reduce the
- 15 proliferation rate or survival rate of the cells having only the targeted allelic form to no more than one half of the proliferation rate or survival rate of cells having at least one non-targeted allelic form. More preferably, the fraction is no more than 1/5 or 1/10, and still more preferably no more than 1/20, 1/50, 1/100, or even lower.
- 20 In a related aspect, the invention provides inhibitors potentially useful for tumor, *e.g.*, cancer treatment, or treatment of other proliferative disorders. Such inhibitors are active on a specific allele of a gene which has at least two different alleles encoding an essential gene product in one of the target gene categories above. Such inhibitors can, for example, be identified by the above screening
- 25 methods.

In a related aspect, the invention provides methods for producing inhibitors active on such specific allelic forms of belonging to one of the above categories genes by

identifying a gene encoding an essential gene product which has alternative allelic forms in a non-tumor cell and which undergoes LOH in a tumor cell, screening to identify an inhibitor which is active on at least one but less than all of the alleles of the gene, and synthesizing the inhibitor in an amount sufficient to produce a  
5 therapeutic effect when administered to a patient suffering from a tumor in which tumor cells have only the allele on which the inhibitor is active.

In the context of this invention, the term "active on an allelic form" or "allele specific inhibitor" or "specific for an allelic form" indicates that the relevant inhibitor inhibits an allele having a particular sequence to a greater extent  
10 (preferably  $\geq 2x$ ) than an allele having a sequence which differs in a particular manner. Thus, for alleles for which a particular base position is identified, the inhibitor has a higher degree of inhibition when a certain base is in the specified position than when at least one different base is in that position. This means that for substitution at a particular base position, at least two of the possible allelic  
15 forms differ in sensitivity to an inhibitor. Usually, however, for a specific sequence variance site, the site will be occupied by one of only two bases. Further, if an inhibitor acts at the polypeptide level, and any of three bases may be present at a particular position in a coding sequence but only one of the substitutions results in an amino acid change, then the activity of the inhibitor  
20 would be expected to be the same for the two forms producing the same amino acid sequence but different for the form having the different amino acid sequence. Other types of examples can also occur.

The term "less active" indicates that the inhibitor will inhibit growth of or kill a cell containing only the allelic form of a gene on which the inhibitor is more active  
25 at concentrations at which it does not significantly inhibit the growth of or kill a cell containing only an allelic form on which the inhibitor is less active.

The term "drug" or "inhibitor" refers to a compound or molecule which, when brought into contact with a gene, its RNA transcript, or its gene product which the compound inhibits, reduces the rate of a cellular process, reduces the level of a cellular constituent, or reduces the level of activity of a cellular component or process. This description is meant to be illustrative of the understanding of the meaning of the term to those skilled in the art and not limiting. Thus, the term generally indicates that a compound has an inhibitory effect on a cell or process, as understood by those skilled in the art. Examples of inhibitory effects are a reduction in expression of a gene product, reduction in the rate of catalytic activity of an enzyme, and reduction in the rate of formation or the amount of an essential cellular component. The blocking or reduction need not be complete, in most cases, for the inhibitor to have useful activity. Thus, in the present invention, "inhibitors" are targeted to genes, their RNA transcript, or their protein product that are essential for cell viability or proliferation. Such inhibitors would have the effect of inhibiting essential functions, leading to loss of cell viability or inhibition of cell proliferation. In preferred embodiments, such inhibitors cause cell death or stop cell proliferation. In preferred embodiments of this invention, inhibitors specifically include a molecule or compound capable of inhibiting one or more, but not all, alleles of genes, their RNA transcript, or their protein product that are essential for cell survival or proliferation. The terms "inhibitor of a gene" or "inhibitor of an allele" as used herein include inhibitors acting on the level of the gene, its gene product, its RNA transcript, its protein product, or modifications thereof and is explicitly not limited to those inhibitors or drugs that work on the gene sequence itself.

Several types of inhibitors are generally recognized in the art. A "competitive" inhibitor is one that binds to the same site on the gene, its RNA transcript or gene product as a natural substrate or cofactor that is required for the action of the gene or gene product, and competitively prevents the binding of that substrate. An

- “allosteric” inhibitor is one that binds to a gene or gene product and alters the activity of the gene or gene product without preventing binding of a substrate or cofactor. Inhibition can also involve reducing the amount of the gene, RNA transcript, or its protein product, and thus the total amount of activity from the gene in the cell. Such inhibition can occur by action at any of a large number of different process points, including for example by inhibiting transcription or translation, or by inducing the elimination of the gene, its RNA transcript, or its protein product where elimination may involve either degradation of the target or egress or export from the compartment in which it is active and the process of excretion or export. Inhibition can also be achieved by modifying the structure of the target, interfering with secondary modifications, or interfering with cofactors or other ancillary components which are required for its activity. Inhibitors can be comprised of small molecules or polymeric organic compounds including oligopeptides or oligonucleotides.
- 15 The term “active on a gene” or “targeted to a gene” indicates that an inhibitor exerts its inhibitory effect in a manner which is preferentially linked with the characteristic properties of a gene, its RNA transcript or its gene product. Such properties include, for example, the nucleotide sequence of the gene or transcribed RNA, the amino acid sequence or post-translational modifications of the protein product, the structural conformation of a protein, or the configuration of a protein or RNA with other cellular constituents (RNA, protein, cofactors, substrates, etc.) required for activity. Thus, in general these terms indicate that the inhibitor acts on the gene, its RNA transcript, its protein product, its gene product, or modifications thereof, or on a reaction or reaction pathway necessarily involving such a gene product to a greater extent than on genes or gene products generally.

A “reduction of the level of activity” of a gene product or allele product refers to a decrease in the functional activity provided by that product. This can be due to

any of a variety of direct causes, including for example, a reduction in the amount of a biologically active molecule present, a change in the structure or modifications of normally active molecules to produce inactive or less active molecules, blockage of a reaction in which the product participates, and blockage  
5 of a reaction pathway in which the product necessarily participates.

In another related aspect the invention provides methods for treating a patient suffering from a proliferative disorder in which an essential gene from one of the above categories has undergone loss of heterozygosity. The method involves administering a therapeutic amount of an allele specific inhibitor of such an  
10 essential gene to a patient whose normal somatic cells are heterozygous for that gene but whose tumor cells contain only a single allelic form of the gene. The inhibitor is active on the specific allele of the gene present in the tumor cells.

A "therapeutic effect" results, to some extent, in a measurable response in the treated disease or condition. Thus, a therapeutic effect can include a cure, or a  
15 lessening of the growth rate or size of a lesion such as a tumor, or an increase in the survival time of treated patients compared to controls, among other possible effects.

The term "therapeutic amount" means an amount which, when administered to a mammal, *e.g.*, a human, suffering from a disease or condition, produces a  
20 therapeutic effect.

In preferred embodiments of this treatment method, the method also involves determining whether the normal cells of the patient are heterozygous for the particular essential gene and determining whether tumor cells of the patient contain only a single allelic form of that gene. The determining may be performed  
25 on a variety of normal cells, such as blood or normal tissue, and on tumor cells.

Either or both of the normal cells and tumor cells may be cultured prior to the determination. The determination may also be carried out using cells retrieved from a frozen or preserved tissue specimen, *e.g.*, from pathological specimens of a patient's tumor and/or normal tissue preserved in a pathology laboratory. Also,  
5 the determining may be performed using a variety of techniques, which may, for example include one of more of: hybridization with an allele specific oligonucleotide probe, hybridization to a gridded set of oligonucleotides, restriction fragment length polymorphism, denaturing gradient gel electrophoresis, heteroduplex analysis, single strand conformation polymorphism, ligase chain  
10 reaction, nucleotide sequencing, primer extension, dye quenching, sequence specific enzymatic or chemical cleavage, mass spectroscopy, and other methods known in the art.

In a related aspect, the invention provides a method for preventing the development of cancer. The method involves administering to a patient having a  
15 precancerous condition or an early stage cancer or cancers an allele specific inhibitor targeted to an allele of an essential gene for which the normal somatic cells of the patient are heterozygous and which has undergone LOH in cells involved in the precancerous condition. In a case where the cells of the precancerous condition are not clonal from a single cell, the method involves  
20 subsequently administering to the patient a second allele specific inhibitor in an amount sufficient to inhibit and preferably kill cells with LOH in which an allele not targeted by the first inhibitor is the only remaining allele of the gene. In most cases, the second allele specific inhibitor will target the alternative allele of the gene targeted by the first inhibitor. However, the second inhibitor can also target  
25 an allele of a second essential gene which has undergone LOH. The second gene may have undergone LOH in the same deletion that affected the first gene due to their proximity on a chromosome, though this is not essential. Additionally, in other cases, allele specific inhibition of one of the alleles of each of 3, 4, or even



more target genes can be utilized in a serial manner (where the patient is heterozygous for each targeted gene). In this case the different target genes need not be tightly linked so that LOH of the various genes does not necessarily occur together. By using the serial inhibition of an allele of each of the target genes, it is possible to inhibit and preferably kill the full population of precancerous cells in which LOH has occurred. Thus, the net effect is essentially the same as if allele specific inhibitors of each of the two alternative alleles of one essential gene had been used.

In the context of the administration of multiple allele specific inhibitors, the terms “serial” or “subsequently” indicates that the administration of two or more inhibitors is sufficiently temporally separated so that normal somatic cells remain functional and are therefore able to survive and/or proliferate. Those skilled in the art will recognize that the required time will depend on various factors, such as clearance rate, type and extent of the effect of an inhibitor on normal cells, and additive cellular toxicity, and that appropriate timing can be routinely determined for particular selections of compounds.

In another related aspect, the invention provides a method for identifying a potential patient for treatment with an inhibitor active on a specific allele of an essential gene from one of the above categories. The method involves identifying a patient having a proliferative disorder characterized by LOH, *e.g.*, a cancer, whose normal somatic cells are heterozygous for the essential gene and determining whether tumor cells in the patient contain only a single allele of the gene. Thus, if the patient is normally heterozygous and the neoplastic cells contain only a single allele of the gene, then the patient is a potential patient for treatment with the inhibitor.

With respect to identifying patients with precancerous or oligoclonal proliferative

- diseases characterized by LOH, and selecting appropriate allele or variance-specific inhibitors for such patients, in some cases it may not be practical to obtain samples of all proliferative lesions for LOH assays.. For example, atherosclerotic plaques in the aorta cannot routinely be sampled by biopsy, and dysplastic lesions
- 5 in the cervix, colon, or bronchus can be multifocal. Therefore, allele specific inhibitors can be selected for such conditions based on previously established patterns of LOH for the condition, and on specific testing for heterozygosity in a given patient. Characteristic patterns of LOH involving specific chromosomes or chromosomal regions have been reported in the art (by Vogelstein's group and
- 10 others) for premalignant changes in the colon, such as adenomatous polyps, polyps with dysplasia and polyps with carcinoma *in situ* (pre-invasive cancer) (Fearon, E. and B. Vogelstein). These studies demonstrate LOH on chromosomes 5q, 17p, and 18q in the earliest lesions. Similar studies have been performed for other premalignant conditions. It will be evident to one skilled in the art that similar
- 15 studies can be readily performed on other conditions characterized by LOH using retrospective analysis of tissue from pathological specimens. The optimal regions for allele or variance specific targeting will be those which are affected by LOH in a high fraction of lesions and in a high fraction of patients. Preferably, at least 40% of lesions will have LOH for a specific target gene, more preferably 60, 80,
- 20 or 90%, and most preferably 100%. However, it is not necessary that 100% of lesions show LOH for a successful treatment by allele specific inhibitors because 2,3,4, or even more inhibitors can be used in a combined approach to target an ever higher fraction of lesions, and because substantial therapeutic benefit may be achieved by inhibiting the proliferation of less than 100% of lesions.
- 25 In a related aspect, the invention provides a method for treating a patient having a proliferative disorder, *e.g.*, suffering from a cancer. The patient's normal somatic cells are heterozygous for an essential gene from one of the above categories, but the patient's cancer cells, or other abnormally proliferating cells,

have only a single allelic form of the gene. This method combines the identification and treatment methods described in the preceding aspects.

- In another aspect, the invention provides a method for identifying a potential patient undergoing transplantation for treatment with an inhibitor active on a
- 5 specific allele of an essential gene from one of the above categories. The method involves identifying a patient undergoing an allogenic transplantation in which the tissue of the donor contains at least one form of an essential gene that is different from those of the recipient. In a preferred aspect of this invention the donor or recipient is homozygous for an alternative form of an essential gene that differs
- 10 from those present in the other. The term "homozygous" means that the two alleles of a gene present in somatic cells contain the same allele or alleles with identical sequence at at least one variant position that determines the activity of an allele specific drug. Such identification then allows methods of treating such patients by targeting the differing variances or allelic forms.
- 15 The term "allogenic" transplantation refers to transplantation of a tissue or cell from the same species which contains different surface antigens than the recipient. In contrast, an "autologous" transplantation is one in which the patient receives their own tissues (commonly bone marrow) that contain identical surface antigens. The surface antigens are commonly those referred to as "histocompatibility" antigens
- 20 or "HLA" antigens which allow the immune system to recognize the patient's own tissues from foreign tissue. In an allogenic transplant, the antigens on the donor tissue are different from those of the recipient. This can lead to an immune response in which the antigens on the transplanted tissue stimulate the patient's immune system to destroy or reject the transplanted tissue. Alternatively, in bone
- 25 marrow transplantation, the antigens on the patient's normal tissue can stimulate the immune system constituted from the donor tissue to destroy the patient's normal tissues. This is termed "graft versus host disease" (GVH).

In a related aspect, the invention provides a method for treating graft versus host disease in allogenic transplantation in which an allele specific inhibitor is used to inhibit proliferation of donor cells, *e.g.*, to inhibit stimulation of the donor immune system. In preferred embodiments, the allele specific inhibitor is selected  
5 by identifying alternative variances or allelic forms of an essential gene that are present in the donor tissues but not the recipient. Therapy with a variance or allele specific inhibitor or inhibitors that recognizes both alleles of the essential gene that are present in the donor, but not both alleles of the same gene that are present in the recipient, can be used to suppress the immune response against the  
10 patient's tissues (GVH) without toxicity to these tissues. Most commonly, the donor tissue would be homozygous for a variance in the essential gene and the recipient would be homozygous to an alternative nucleotide or amino acid at a specificity determining site of variance. However, alternative combinations can also be used which result in at least one allelic form being present in the recipient  
15 which is not present in the donor cells, for example the donor could be homozygous and the recipient could be heterozygous for different allelic forms. As in other aspects described, a plurality of target genes can also be utilized.

In another aspect, the invention provides a method for enhancing engraftment of an allogenic bone marrow transplant in which an allele specific inhibitor is used  
20 to kill or suppress the patient's own bone marrow, providing "space" for engraftment of the donor cells within the marrow cavity. In preferred embodiments, the allele specific inhibitor is selected by identifying alternative forms of an essential gene that are present in the recipient but not the donor marrow. Therapy with an allele specific (generally a variance specific) inhibitor  
25 that recognizes both forms of the essential gene that are present in the recipient, but not both forms of the same gene that are present in the recipient, can be used to suppress the patient's own marrow without toxicity to the transplanted cells. It will be recognized by those in the art that this method can be used to reduce the

frequency of chimerism and increase the rate of success in engrafting an allogenic marrow.

- “Chimerism” refers to a transplantation that is incomplete, leading to the proliferation of bone marrow progenitor cells derived from both the donor and recipient. Chimerism is generally an undesirable outcome that commonly results in gradual elimination of the graft due to competition with the patient’s own cells. Allele specific inhibitors can be used to treat or prevent chimerism by selectively killing or suppressing proliferation of the patient’s own cells without toxicity to the donor cells.
- 10 In another aspect, the invention provides a method for treating cancer in a patient receiving allogenic or autologous transplantation in which an allele specific inhibitor is used to kill or inhibit the growth of cancer cells without toxicity to the transplanted marrow. In one embodiment, in an autologous transplantation the allele specific inhibitor is selected to recognize one alternative allele of an essential
- 15 gene remaining in the cancer cell due to LOH in patients who are heterozygous with two different alternative forms of the essential gene in their normal cells and in the autologous bone marrow graft. Treatment with such a drug will enable continuing therapy of cancer without suppression of the transplanted marrow. In an alternative embodiment, in an allogenic transplantation, therapy with an allele
- 20 specific inhibitor that recognizes the one form of the essential gene that is present in cancer cells due to LOH in the recipient, but not an alternative form or forms of the same gene that are present in the recipient’s normal cells and in the donor cells can be used to treat the cancer in the patient without toxicity to the transplanted cells. It will be recognized by those in the art that such therapy will enable more
- 25 effective cancer therapy during and after transplantation. Moreover, such therapy would preserve the function of the immune system which is an important element in effective cancer therapy.

In a related aspect, the invention can be used *ex vivo* during autologous transplantation to eliminate malignant cells from the transplanted marrow. The principle of autologous bone marrow transplantation is that bone marrow can be harvested from a patient prior to high dose radiation or chemotherapy that would normally be lethal to the bone marrow. Following such therapy, the patient can then be treated by reimplantation of their own marrow cells to reconstitute the bone marrow and hematopoietic functions. An important limitation of this procedure is that bone marrow harvested prior to such therapy often contains many malignant cells, and that implantation of the harvested bone marrow often results in reseeding of the patient's malignancy. Various techniques for "purging" the bone marrow of such malignant cells have been described. These methods are focused on selecting "normal" bone marrow stem cells or progenitor cells that are within the harvested tissue for selective reimplantation. The present invention provides for an improved method for purging bone marrow of malignant cells using allele specific inhibitors of essential genes. The method involves identifying an essential gene with only one variant form remaining in the cancer cells due to LOH in patients who are heterozygous with two different alternative forms of the essential gene in their normal cells (and in the autologous bone marrow). The patient's bone marrow is then cultivated *ex vivo* using methods known in the art in the presence of an allele specific inhibitor that inhibits the allele that is present in the cancer cells, but not the alternative allele that is present in the heterozygous normal bone marrow. This treatment will result in killing of cancer cells within the graft, enabling selective reimplantation of normal cells. It will be recognized that one or more drugs could be used simultaneously or sequentially in this manner to achieve more efficient purging of cancer cells.

In another aspect, the present invention provides a method for sorting cells, for example for separating cancer cells from normal cells during an autologous bone marrow transplantation. The method utilizes a compound, preferably an antibody or

antibody fragment, which specifically binds to at least one but less than all the products of alleles which occur in a population of a particular gene which encodes a cell surface protein. Such a binding compound is used to bind with cells which express a targeted allele. If cancer cells from a patient who is heterozygous for that gene (having both a targeted allele and a non-targeted allele) have undergone LOH of the particular gene such that only the non-targeted allele is present in the cancer cells, then the binding compound can be used to bind to normal cells and to pull them out from a mixture of normal and cancer cells. This separation is possible because the binding compound will bind to the protein from the targeted allele of the gene expressed in the normal cells, but will not recognize and will not bind to the cancer cells as there is no product of the targeted allele present on those cells. Use of this method thus allows the isolation of normal cells, which can then be reintroduced to the marrow in an autologous transplant following anticancer treatment of the patient, thereby avoiding the problem of reintroduction of cancer cells. In this method, the targeted gene need not be an essential gene, or have any particular function. All that is needed is that the gene product be accessible or can be made accessible to the allele specific binding compound and that there be alternative allelic forms of the gene present such that the products can be distinguished by allele specific binding compounds and that the gene have undergone LOH between the normal cells and the cancer cells. However, it is also recognized that this method can also be used to separate any sets of cells which express different allelic forms of a gene where the gene products are accessible to allele specific binding compounds.

In preferred embodiments, the binding compound is immobilized, such as on a solid support, or can be caused to leave solution, such as by precipitation or by sandwich binding of the binding compound with a second binding compound, so that the bound cells are directly removed from the mixture. In other embodiments, the binding compound allows the recognition of the targeted cell, such that the cells can

be separated mechanically, for example using fluorescence activated cell sorting (FACS), or other cell sorting method as known to those skilled in the art. Also in preferred embodiments, the binding compound is an antibody or antibody fragment which retains allele specific binding. Such antibodies can be readily obtained by  
5 conventional methods as polyclonal or monoclonal antibodies after isolation of an appropriate antigen.

In another aspect, the invention provides a method for inhibiting growth of or killing a cell containing only one allelic form of a gene by contacting the cell with an inhibitor active on that allelic form. The gene has at least two sequence  
10 variants in a population, and belongs to one of the categories of essential genes described below. The inhibitor is less active on at least one other allelic form of the gene.

In preferred embodiments of the above aspects in which an allele specific inhibitor is used to inhibit a cell or to treat a patient, a plurality of different inhibitors may  
15 be used. Preferably different inhibitors target a plurality of different variances in a single target gene, or target variances in different target genes, or both. In particular embodiments a plurality of inhibitors is used simultaneously, in others there is serial administration using different inhibitors or different sets of inhibitors in separate administrations, which may be performed as a single set of  
20 administrations in which each set of inhibitors is administered once, or in multiple serial administrations in which each set of inhibitors is administered more than once. Such use of multiple inhibitors provides enhanced inhibition, which preferably includes killing, of the targeted cells. In addition, allele specific inhibitors as described can be used in conjunction with other treatments for  
25 diseases and conditions, including in conjunction with other chemotherapeutic agents such as other antineoplastic agents.



In a related aspect, an allele specific inhibitor can be used in conjunction with a conventional antiproliferative or chemotherapeutic agent or therapy, such therapies including radiation, immunotherapy, or surgery. In preferred embodiments the conventional therapy causes one or more genes within the cancer cell, or  
5 noncancer proliferative lesion, to be essential for cell survival that are would not be essential in the absence of said conventional therapy. For example, the treatment of cancer with radiation or alkylating agents makes efficient DNA repair essential for cell survival. In another example, depleting cancer cells of certain nutrients may make certain synthetic metabolic pathways essential. These  
10 examples are meant to be illustrative of the use of the present invention to those skilled in the art and not limiting. Further discussion and examples of the use of conditionally essential genes and their utilization in the methods of this invention are provided in the Detailed Description and the Examples.

In accord with the above aspects, in a further aspect the invention provides a  
15 pharmaceutical composition which includes at least one allele specific inhibitor. In preferred embodiments the composition includes at least one allele specific inhibitor and a pharmaceutically acceptable carrier. Such carriers are known in the art and some commonly used carriers are described in the Detailed Description below. Also in preferred embodiments the composition includes two, three, or  
20 more allele specific inhibitors, and may also include a pharmaceutically acceptable carrier. In other preferred embodiments, the composition includes at least one allele specific inhibitor and another antineoplastic agent, which need not be an allele specific inhibitor. The embodiments of this aspect may also optionally include diluents and /or other components as are commonly used in  
25 pharmaceutical compositions or formulations. In embodiments having a plurality of allele specific inhibitors, the inhibitors may target a plurality of different variances of a single target essential gene, or may target sequence variances of a plurality of different essential genes or combinations thereof.

In accord with the use of pharmaceutical compositions, the present invention also provides a packaged pharmaceutical composition comprising an allele specific inhibitor as described above, bearing a Food and Drug Administration use indication for administration to a patient suffering from a cancer or suffering from  
5 another proliferative disorder.

Determinations of essential gene heterozygosity and tumor cell LOH may be performed by a variety of methods, such as direct sequencing of known sequence variance sites and probe hybridization with variance specific probes. Thus, the invention also provides a nucleic acid probe at least 9, 12, 15 or 20 nucleotides in  
10 length, but preferably not more than 30 nucleotides, which will hybridize to a portion of a first allelic form of an essential gene in one of the above categories under specified hybridization conditions and not to a second allelic form under those hybridization conditions, the first and second allelic forms have a sequence variance within the complementary sequence. Preferably the probe is at least 12  
15 nucleotides in length and is perfectly complementary to a portion of the first allelic form which includes a sequence variance site. The probe hybridizes under stringent hybridization conditions to the portion of the first allelic form and not to the corresponding portion of the second allelic form. This means that the probe does not bind to the second allelic form to an extent which prevents identification  
20 of the preferential specific binding to the first allelic form. The thermodynamics of the probe hybridization can be predicted to maximize the desired differential hybridization, providing optimization for probe length, sequence, structural modifications, and modifications to hybridization conditions.

The invention also provides nucleic acid probes or primers adjacent to the site of a  
25 variance that can be used to amplify a sequence containing the variant position to determine which variance is present at that position. Such probes or primers can readily be designed based on the sequences provided in the corresponding database

sequence entry or otherwise determined. The method of determining the variance can involve allele specific hybridization, sequencing or analysis of the amplified fragment by mass spectroscopy, SSCP, gene sequence database analysis, capillary electrophoresis, bindase/resolvase systems, or other methods known in the art. In  
5 a preferred embodiment, the amplified sequence spans more than one variant position and the method used for determining the variances identifies which variances are present at each position and combinations of variances that are present on each allele.

In preferred embodiments of the above aspects, the specific target allelic form has  
10 the characteristics as described above. Thus, for aspects in which the category of gene is specified, in preferred embodiments the gene belongs to a particular sub-category, for example, subcategories as specified in Table 1. Also in preferred embodiments, the gene is an identified target gene as listed in Table 1 or otherwise specified herein, including targeting utilizing the specified variances for exemplary  
15 genes described herein, singly or in combination in an allelic form. Also in preferred embodiments, the target gene is an allelic form having characteristics as specified above, for example is a gene which has a high frequency of heterozygosity and/or occurs in a chromosomal region which undergoes LOH in a cancer at a frequency as specified above. For aspects in which the target gene has  
20 a specified LOH frequency, the LOH frequency may be provided by published literature, inferred from the LOH of nearby genetic members, or independently determined, such as by the methods known in the art.

The use of conditionally essential genes for a number of applications is similar to the aspects above, but generally also involve an alteration of environment to make  
25 the gene essential and also provides additional aspects. For a conditionally essential gene, the essentiality may, but need not be absolute. Instead, in this context, the term "essential" means that the gene confers a significant advantage,

such that the growth or survival of the non-targeted cells is preferably at least 2x, more preferably 3x, 4x, 5x, 10x, or more as compared to the targeted cells.

Thus, similar to the above, the invention provides a method for identifying an inhibitor potentially useful for treatment of cancer or other proliferative disorder.

- 5 The inhibitor is active on a conditionally essential gene, and the gene is subject to loss of heterozygosity in a cancer. The method includes identifying at least two alleles of a said gene which differ at at least one sequence variance site and testing a potential allele specific inhibitor to determine whether the potential inhibitor is active on at least one but less than all of the identified alleles. If the potential
- 10 inhibitor inhibits expression of at least one but less than all of the alleles or reduces the level of activity of a product of at least one but less than all of the alleles, this indicates that the potential allele specific inhibitor is, in fact such an allele-specific inhibitor inhibitor.

- In preferred embodiments of this and the various aspects described below, the
- 15 conditionally essential gene is one of the exemplary genes presented in the table of conditionally essential genes or in the examples.

- Similar to other types of target genes described above, the invention provides inhibitors, methods for producing inhibitors, pharmaceutical compositions, methods for identifying potential patients, probes, and primers which target or recognize
- 20 alleles of a conditionally essential gene or utilize inhibitors which target such genes.

The invention also provides methods for preventing the development of cancer, methods for treating a patient suffering from a cancer, and methods for inhibiting growth of a cells as described above except that the targeted cells are subjected to an altered condition such that the gene becomes essential.

In still another aspect, not requiring the use of allele specific inhibitors, but still utilizing information about sequence variance or allelic differences between normal somatic cells and cancer cells in a patient, the invention provides a method for selecting a patient for treatment with an antiproliferative treatment. The method

5 includes the following steps: determining whether normal somatic cells in a potential patient are heterozygous for an essential or conditionally essential gene, where a first allelic form of the gene is more active than a second allelic form, and where a reduction in the activity of the gene in a cell increases the sensitivity of that cell to an antiproliferative treatment; and determining whether cancer cells from the

10 patient have only the second allelic form of the gene. If the somatic cells are heterozygous and the cancer cells have only the second allelic form, this indicates that the patient is suitable for treatment with the antiproliferative treatment because the cancer cells will be more sensitive to the antiproliferative treatment. In preferred embodiments, the antiproliferative treatment is radiation or administration of a

15 cytotoxic drug.

In a related aspect, the differences between the normal somatic cells and the cancer cells in a patient are used in a method for selecting an antiproliferative treatment for a patient suffering from a cancer. This method involves determining whether there will be a differential effect of the prospective treatment on the cancer cells as

20 compared to the normal cells based on a differential response of the cancer cells due the presence in the cancer cells of only the less active form of a conditionally essential gene which is present in two alternative allelic forms with differing activities in the somatic cells. The method thus involves determining whether normal somatic cells in a potential patient are heterozygous for an essential or

25 conditionally essential gene which reduces the sensitivity of cells to an antiproliferative treatment. As noted, a first allelic form of the gene is more active than a second allelic form, and a reduction in the activity of the gene in a cell increases the sensitivity of that cell to the prospective antiproliferative treatment;

and determining whether cancer cells of said patient have only the second, less active, allelic form of the gene. If these factors are present, this indicates that the proposed treatment is suitable for that patient.

- In preferred embodiments of above aspects, a conventional therapy acts on a protein or other molecular target in the same pathway as the allele specific inhibitor. As an example, the antineoplastic drug hydroxyurea, which inhibits ribonucleotide reductase (RR), can be used in conjunction with an allele specific inhibitor of RR subunit M1 or M2 or another gene that encodes a product important in nucleotide synthesis. Similarly, the antiproliferative drug methotrexate inhibits the enzyme dihydrofolate reductase (DHFR), and can be used with allele specific inhibitors of DHFR that would result in a differential methotrexate effect on cancer tissues compared to normal proliferating tissues. Alternatively, methotrexate can be used with allele specific inhibitors of other genes important in folate metabolism to achieve an enhanced cancer cell specificity for methotrexate. Similarly, the anticancer drug 5-fluorouracil and related compounds can be administered together with an allele specific inhibitor of thymidylate synthase (TS) in a patient heterozygous for TS and with LOH at the TS gene in proliferating cells, e.g., cancer cells. Alternatively, an allele specific inhibitor of 5-FU degradation or metabolism can be administered with 5-FU. For example, the enzyme dihydropyrimidine dehydrogenase, which catalyzes the first and rate limiting step in 5-FU catabolism would have the effect of potentiating 5-FU action in cancer cells due to their lesser ability to metabolically inactivate 5-FU. One skilled in the art will readily recognize that similar methods can be used with other conditionally essential genes, including specific genes listed in the table of conditionally essential genes.
- Some conditionally essential genes occur in active and less active, or nearly inactive allelic forms. Further, some cancer patients are heterozygous for active and less active forms in their normal tissues, but due to LOH, their cancer cells contain only

the less active allelic form. As describe above, such patients can be identified by a diagnostic test of their normal cells and cancer cells. Such a test will identify which patients should be treated with a specific treatment, such as a particular drug or radiation treatment or other treatment. Such a therapy, which is not allele specific, would nonetheless have cancer specific effects due to the LOH-determined difference in the ability of the cancer cells to respond to the cytotoxic or cytostatic effects of therapy.

For example, patients with Ataxia Telangiectasia are homozygous for mutant alleles of the ATM gene. Such individuals are hypersensitive to radiation therapy or radiomimetic drugs. Heterozygotes for normal and mutant ATM are normal and have been estimated to account for 0.5-1% of the North American population, but, due to an increased risk of cancer, may account for up to 5% of some cancers, for example, breast cancer. The ATM gene maps to chromosome 11q23, a region frequently affected by LOH in breast and other cancers. In breast cancers arising in ATM heterozygotes in which the more active (normal) ATM allele is lost in cancer tissue due to LOH, treatment with radiation or radiomimetic drugs would be differentially toxic to cancer cells. It has been shown that ATM heterozygotes are less sensitive to such treatments than ATM mutant (less active) homozygotes. Such use of an LOH diagnostic procedure to select appropriate antineoplastic therapy represents a change from the current procedures which are based solely on tissue origin, grade, and stage of cancer.

In such an approach, preferably the difference in activity between more active and less active allelic forms is at least 2x, more preferably at least 3x, 4x, or 5x, and most preferably at least 6x, 10x, or even more.

Preferably a target conditionally essential gene is one such that at least 0.1%, 0.5%, 1% or 5%, or the higher rates as stated above, of a population is

heterozygous for a particular sequence variance

Additional specific genes within the categories or subcategories described which are potentially useful for allele specific therapy can be readily identified by those skilled in the art using the methods described herein and/or using information  
5 available to those familiar with cellular genetics and tumor biology. In particular such genes can be identified and/or obtained by identifying essential genes, determining whether the gene contains sequence variants in a population, determining whether the gene undergoes LOH in one or more tumors or other proliferative disorders. Genes having these characteristics can then be used for  
10 identifying allele specific inhibitors and evaluated for use in the other methods of this invention. Such procedures are routine, as is shown by the Detailed Description of the Preferred Embodiments below, including the Examples.

In preferred embodiments of the above methods and inhibitors involving particular target genes or classes or categories of genes, the inhibitor or potential inhibitor is  
15 a ribozyme which is designed to specifically cleave a particular target allelic form of a gene (*i.e.*, a nucleotide sequence such as mRNA).

The ribozyme is designed to cleave the nucleotide (*e.g.*, RNA) sequence at a position in the nucleotide chain of the target allelic form at or near the position of a sequence variance. Usually the ribozyme will have a binding sequence which is  
20 perfectly complementary to a target sequence surrounding the sequence variance site. Preferably, the ribozyme does not consist of only ribonucleotides, and therefore includes at least one nucleotide analog or modified linkage. In preferred embodiments the ribozyme has a hammerhead or hairpin motif, but may have other structural motifs as known to those skilled in the art..

25 The term "ribozyme" refers to a catalytic RNA molecule, including those



commonly referred to as hammerhead ribozymes and hairpin ribozymes, generally having an endonuclease activity, but includes catalytic RNA molecules, catalytic DNA molecules (DNAzymes), and derivatives of such molecules unless indicated to the contrary. In particular, as understood by those skilled in the art, ribozymes  
5 may incorporate a variety of nucleotide analogs, modified linkages, and other modifications.

In connection with ribozymes, "target sequence" refers to a nucleotide sequence which includes a binding site and a cleavage site for a ribozyme. For use in this invention, preferably a gene having a ribozyme target sequence exists in two  
10 allelic forms in normal somatic cells of a patient. The two allelic forms differ in nucleotide sequence within the target sequence, *i.e.*, have a sequence variance within the target sequence.

Also in connection with ribozymes, the term "specifically cleaves" means that a particular ribozyme will cleave a target sequence to a greater extent than it will  
15 cleave a different sequence. For allele specific ribozymes, this means that for two allelic forms having a sequence variance in the target sequence, preferably the ribozyme will cleave one of the allelic forms more efficiently than the other. Those skilled in the art will understand that the target discrimination can be provided by base differences within the ribozyme binding sequence of the  
20 substrate at or close to the cleavage site.

Similarly, in preferred embodiments the inhibitor or potential inhibitor is an oligonucleotide, e.g., an antisense oligonucleotide, preferably at least partially an oligodeoxyribonucleotide. The antisense oligonucleotide is complementary to a sequence which includes a sequence variance site. Usually, though not  
25 necessarily, the antisense oligonucleotide is perfectly complementary to a sequence of the target allelic form which includes a sequence variance site. The antisense

oligonucleotide preferably is at least twelve nucleotides, more preferably at least seventeen nucleotides in length. In some cases the antisense oligonucleotide may advantageously be longer, for example, at least 20, 25, or 30 nucleotides in length. Also in preferred embodiments, the oligonucleotide is no longer than 20, 5 25, 30, 35, 40, or 50 nucleotides. The optimal length will depend on a number of factors, which may include the differences in binding free energy of the oligonucleotide to the target sequence as compared to binding to the non-target allelic form, *i.e.*, the non-target sequence variant, or the kinetics of nucleic acid hybridization. The oligonucleotide preferably contains at least one nucleic acid 10 analog or modified linkage. Such complementary oligonucleotides may function in various ways, and those skilled in the art will know how to design the oligonucleotide accordingly. Such functional mechanisms include, but are not limited to direct blocking of transcription of a gene by binding to DNA (*e.g.*, high affinity antisense, including triple helix), direct blocking of translation by binding 15 to mRNA, RNaseH mediated cleavage of RNA or other RNAase mediated cleavage, and binding-induced conformational changes which block transcription or translation or alter the half-life of mRNA. Triple-helix modes of action include the formation of a triple-helical structure between the two strands of genomic DNA and an antisense molecule, *i.e.*, anti-gene strategy, or between an RNA 20 molecule and an antisense oligonucleotide which loops back to contribute two of the three strands of the triple helix, or between an RNA and an antisense where the RNA provides two of the three strands of the triple helix.

The term "oligonucleotide" refers to a chain molecule comprising a plurality of covalently linked nucleotides as recognized in the art. The oligonucleotide 25 preferably has about 200 or fewer backbone units corresponding to nucleotide subunits, more preferably about 100 or fewer, still more preferably about 80 or fewer, and most preferably about 50 or fewer. An oligonucleotide may be modified to produce an oligonucleotide derivative. Unless indicated otherwise the

term "oligonucleotide" includes "oligonucleotide derivatives".

A large number of nucleic acid modifications are known in the art which may be used in the nucleic acid molecules of the present invention, thereby producing "nucleic acid derivatives" or "oligonucleotide derivatives". Such modifications  
5 can be used, for example, to enhance resistance to degradation by nucleases or to modify functional characteristics such as binding affinity. In preferred embodiments, the ribozyme, antisense oligonucleotide, or other nucleic acid molecule contains at least one modified linkage, including but not limited to phosphorothioate, phosphoramidate, methylphosphonate, morpholino-carbamate,  
10 and terminal 5'-5' or 3'-3' linkages. Also in preferred embodiments, the nucleic acid molecule contains at least one nucleotide analog. Such analogs include but are not limited to nucleotides modified at the 2' position of the ribose sugar, *e.g.*, 2'-O-alkyl (*e.g.*, 2'-O-methyl or 2'-methoxyethoxy) or allyl, 2'-halo, and 2'-amino substitutions, and/or on the base (*e.g.*, C-5 propyne pyrimidines), and  
15 analogs which do not contain a purine or pyrimidine base, and includes the use of nucleotide analogs at the terminal positions of a nucleic acid molecule. Preferably a 2'-O-alkyl analog is 2'-O-methyl; preferably a 2'-halo analog is 2'-F.

A specific embodiment of this invention is the use of hybrid oligonucleotides that contain within a linear sequence two different types of oligonucleotide  
20 modifications. In a particular embodiment, these modifications are used such that a segment of the oligonucleotide that hybridizes to the sequence variance is RNAase sensitive, but other segments are not RNAase sensitive.

Other modifications may also be used as are known in the art, such as those described in connection with antisense and triple helix in: Crooke & Bennett,  
25 1996, *Annual Rev. Pharm. and Toxicol.* 36:107-129; Milligan et al., 1993, *J. Med. Chem.* 36:1923-1937; Reynolds et al., 1994, *Proc. Nat. Acad. Sci. USA*

91:12433-12437; and McShan et al., 1992, *J. Biol. Chem.* 267-5712-5721, which are hereby incorporated by reference. An additional modification useful for delivery of oligonucleotides is complexation of oligonucleotides with nanoparticles, as described in Schwab et al., 1994, *Proc. Nat. Acad. Sci. USA* 91:10460-10464. As described further below, oligonucleotides may be complexed with other components known in the art which provide protection and/or enhanced delivery for the oligonucleotides, and may be useful for either gene delivery or for delivery of non-coding oligonucleotides.

Thus, "derivatives of nucleic acid inhibitors" include modified nucleic acid molecules which may contain one or more of: one or more nucleotide analogs, including modifications in the sugar and/or the base, or modified linkages, base sequence modifications, and insertions or deletions, or combinations of the preceding. Other derivatives are also included as are known in the art.

Similarly, in preferred embodiments the inhibitor or potential inhibitor is an antibody, preferably a monoclonal antibody, which may be complexed or conjugated with one or more other components, or a fragment or derivative of such an antibody. It is recognized in the art that antibody fragments can be produced by cleavage or expression of nucleic acid sequences encoding shortened antibody molecule chains. Such fragments can be advantageously used due to their smaller size and/or by deletion of sites susceptible to cleavage. In addition, derivatives of antibodies can be produced by modification of the amino acid moieties by replacement or modification. Such modification can, for example, include addition or substitution or modification of a side chain or group. Many modifications and biological effects of such modifications are known to those skilled in the art, and may be used in derivatives of antibodies in accord with those biological effects. Such effects can include, for example, increased resistance to peptidases, modified transport characteristics, and ability to carry a ligand or other

functional moiety. In preferred embodiments, the antibody is a humanized antibody from a non-human animal, *e.g.*, a humanized mouse or rabbit antibody. Many instances of monoclonal antibodies that distinguish protein differing by a single amino acid are known in the art.

- 5 An inhibitor may also be an oligopeptide or oligopeptide derivative. Such peptides may be natural or synthetic amino acid sequences, and may have modifications as described for antibodies above. In general, an oligopeptide will be between about 3 and 50 residues in length, preferably between about 4 and 30, more preferably between about 5 and 20 residues in length.
- 10 In other embodiments, the inhibitor is a small molecule, for example, a molecule of one of the structural types used for conventional anticancer chemotherapy.

- By "small molecule" or "low molecular weight compound" is meant a molecule having a molecular weight of equal to or less than about 5000 daltons, and more preferably equal to or less than about 2000 daltons, and still more preferably equal
- 15 to or less than about 1000 daltons, and most preferably equal to or less than about 600 daltons. In other highly preferred embodiments, the small molecule is still smaller, for example less than about 500, 400, or 300 daltons. As well known in the art, such compounds may be found in compound libraries, combinatorial libraries, natural products libraries, and other similar sources, and may further be
- 20 obtained by chemical modification of compounds found in those libraries, such as by a process of medicinal chemistry as understood by those skilled in the art, which can be used to produce compounds having desired pharmacological properties.

- In connection with the gene sequences or subsequences of gene sequences or
- 25 primer sequences as described herein, the sequences listed under the accession

- number are believed to be correct. However, the genes can be readily identified and the invention practiced even if one or more of the specified sequences contain a small number of sequence errors. The correct sequence can be confirmed by any of a variety of methods. For example, the sequence information provided
- 5 herein and/or published information can be used to design probes for identifying and isolating a corresponding mRNA. The mRNA can be reverse transcribed to provide cDNA, which can be amplified by PCR. The PCR products can then be used for sequencing by standard methods. Alternatively, cDNA or genomic DNA libraries can be screened with probes based on the disclosed or published gene
- 10 sequences to identify corresponding clones. The inserts can then be sequenced as above. If complete sequence accuracy is desired, such accuracy can be provided by redundant sequencing of both DNA strands. Those skilled in the art will recognize that other strategies and variations can also be used to provide the sequence or subsequence for a particular gene.
- 15 Other features and advantages of the invention will be apparent from the following description of the preferred embodiments and from the claims.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

- Fig. 1 shows seventeen gene-specific Target Gene Summary Tables which show variances detected in some of the exemplary genes described as examples in the
- 20 specification. Those genes are:

Sodium, potassium ATPase

CTP synthetase

Ribonucleotide reductase M1 subunit

Thymidylate synthase

- Alanyl tRNA synthetase
- Cysteinyl tRNA synthetase
- Glutamyl-prolyl tRNA synthetase
- Glutaminyl tRNA synthetase
- 5 Lysyl tRNA synthetase
- Threonyl tRNA synthetase
- Ribosomal protein S14
- Eukaryotic initiation factor 5A
- Replication protein A, 70 kD subunit
- 10 Replication protein A, 32 kD subunit
- RNA Polymerase II, 220 kD subunit
- TATA associated factor IIH
- Dihydropyrimidine dehydrogenase

- These tables show, in the title, the name of each gene, its chromosome location
- 15 and the Varia ID number. The horizontal section of the table displays, from left to right, the name of the primers used to amplify the polymorphic segment, the number of the polymorphic nucleotide (the numbering corresponds to the GenBank accession number reported in the central box under 'Sequence from:') and the two alternative sequences at the variant site. Then, under columns 1 - 36, the
- 20 genotypes of 36 lymphoblastoid cell lines are given, followed by the frequency of heterozygotes ('het rate'), a 'Comments' section which describes any unusual aspects of the variances, a 'Location' section which reports the location of any variances and the inferred effect on amino acid sequence, if any, and a 'Race specific heterozygosity' section which reports frequency of heterozygotes in any
- 25 racial groups with particularly high heterozygosity levels. Below the 'Genotypes of 36 unrelated individuals' section the racial or ethnic identity of the subjects is shown (see legend in box at right: 'Ethnic & racial groups surveyed'). The sequence surrounding the variances is shown in the box at bottom left, with the

location of the variant base marked in bold type.

Fig. 2 is a schematic showing the practical flow of the SSCP technique as used for exemplary target genes. This flow chart, in conjunction with the description of the SSCP technique in the Detailed Description, demonstrates how sequence variances of the exemplary genes were identified. In conjunction with published descriptions of the SSCP technique, one skilled in the art can thus readily use SSCP to identify sequence variances in other genes within the scope of this invention.

Fig. 3 is a table describing the extent and distribution of loss of heterozygosity throughout the genome for a number of cancers as reported in the literature. The table is divided into 41 sections, one for each of the chromosomal arms for which there is information about LOH frequency. (There is no information for the short arm [called the p arm] of chromosomes 13, 21 or 22, all of which are very short and contain mostly repetitive DNA.) In each of the 41 sections there is a list of polymorphic loci (sites) that have been tested for LOH in one or more cancer types. The loci are ordered, to the extent that present information allows, from the telomeric end of the short arm of the chromosome to the centromere (p arm tables), or from the centromere to the telomeric end of the long arm of the chromosome (q arm tables). Many chromosomes have not yet been well studied for LOH, so the absence of data on LOH in a particular cancer type on a particular chromosome arm should not be construed as indicating no LOH. It may simply indicate no good LOH studies have yet been published. The Loss of Heterozygosity Table is explained in detail below.

*Column 1* Chromosomes, when stained with dyes such as giemsa, have alternating dark and light staining bands. These bands are the basis of chromosome nomenclature. Many of the markers used for LOH studies have been assigned to



specific chromosome bands, or can be inferred as likely to belong to specific bands based on other information. The 'unknown' notation in this column indicates that the paper from which the data was obtained (column 7) did not provide chromosome band information. In such cases other information has generally been used to order the data, however the order of some markers remains uncertain.

*Column 2* LOH studies are performed with specific DNA markers or probes (for Southern blotting) or with DNA primers (if polymerase chain reaction was used) from a specific site, or locus, on a chromosome. The name of the marker, locus or probe used to perform each LOH assay is given in the second column of the Table, under 'Marker'. In the Table the markers are listed in their likeliest order along the chromosome, from the telomere of the p arm to the centromere for the p arm tables, and from the centromere to the telomere of the q arm for the q arm tables.

*Columns 3, 4 & 5* The total number of cancers evaluable for LOH at the specific marker shown in column 2 (in the paper cited in column 7) are shown in column 3, 'Total'. This is generally the number of patients that were heterozygous for the marker in their normal DNA. Column 4, 'Cases w/LOH', shows the number of patients with LOH at the DNA marker. Column 5, 'LOH Freq', is the quotient of column 4 divided by column 3, giving the fraction of patients with LOH at the indicated marker.

*Column 6* The type of cancer studied is indicated under the heading 'Tumor Type'. In some cases more detailed clinical information on cancer subtype or clinical stage is available in the paper cited in column 7.

*Column 7* The literature citation, or 'Reference', from which the data was drawn. The references are provided in a compact form consisting of journal abbreviation (see the list of journal abbreviations below), volume and page.

**Note**

Studies of allele loss in benign neoplasms or in non-neoplastic conditions are not included in this table.

**Journal Abbreviations for Literature Cited in the Table**

- 5 The abbreviations used in the Tables are as follows:

AJHG = American Journal of Human Genetics

AJP = American Journal of Pathology

B = Blood

BJC = British Journal of Cancer

- 10 C or CA = Cancer

CCG = Cancer Cytogenetics

CGC = Cell Genetics and Cytogenetics

CL = Cancer Letters

CR = Cancer Research

- 15 CSurv = Cancer Surveys

EJC = European Journal of Cancer

G or GE = Genomics

GCC = Genes, Chromosomes & Cancer

GO = Gynecological Oncology

- 20 HG = Human Genetics

HMG = Human Molecular Genetics

IJC = International Journal of Cancer

JAMA = Journal of the American Medical Association

JJCR = Japanese Journal of Cancer Research (Gann)

- 25 JNCI = Journal of the National Cancer Institute

JU = Journal of Urology

Lan = Lancet

LI = Laboratory Investigation

N = Nature

NEJM or NEJ = New England Journal of Medicine

5 O = Oncogene

PN or PNAS = Proceedings of the National Academy of Sciences

S = Science

This data base thus identifies sites and regions of LOH associated with the particular identified cancers, including high frequency LOH chromosomal arms as  
10 well as the identified smaller regions associated with the particular markers. Both as indicated in the Summary and Detailed Description, LOH information such as this identifies essential genes mapping to those LOH regions as likely potential target genes because of the high probability that an essential gene in such a region undergoes LOH at frequencies similar to the marker. Such gene identification  
15 thus further identifies particular cancers which can potentially be treated with inhibitors targeting sequence variances in those essential genes.

The database provided shows information which is contained in published references dealing with cancer LOH. Those skilled in the art will recognize however that similar information can be readily obtained from the published  
20 literature in relation to other cancers and other neoplastic disorders. Thus this table demonstrates that one skilled in the art can readily identify regions of high frequency LOH for other such disorders and cancers, and can further readily identify essential genes which are potential targets for variance specific inhibition and the treatment of the corresponding condition and in other aspects of this  
25 invention.

Fig. 4 is a table summarizing the results in Fig. 3 by chromosome arm. Data for

all loci on each chromosome arm has been summed in a single statistic for LOH frequency on that chromosome arm.

Fig. 5 is a Target Variances by Field Table, which summarizes information on DNA sequence variances in selected genes from the Target Gene Table (Table 1),  
5 and is organized into groups of related genes that parallel the fields in the Target Gene Table.

- The heading at the top of each category of essential genes shows a number and a subcategory name. The number indicates which of the six principal categories of essential genes the subcategory belongs to (e.g. genes required for cell proliferation is category 1, genes required to maintain inorganic ions at levels compatible with cell growth or survival is category 2, etc.).  
10
- Below the heading is a sentence on 'Validation' which briefly refers to some of the data which shows that genes in the subcategory are essential. Summary information on target gene variances is then listed, with five  
15 columns of data.
- The first column gives the Variagenics gene ID number, which serves as a cross reference to the Target Variances Table (see below), where more detailed information on variances can be found.
- The second column lists gene names. (The GenBank accession number in  
20 column 5 may be a more reliable way to identify genes.)
- The third column lists the number of variances found. These variances were detected by a variety of experimental and informatics based procedures described in the examples. Many variances were detected by two independent methods (e.g. informatics based detection and T4 endonuclease  
25 VII detection). A molecular description of the variances is provided in the Target Variances Table (see below).
- The fourth column lists the chromosome location of the target gene, if known. Knowledge of the chromosome location permits assessment of the

cancers in which LOH would be expected to affect the target gene. (See the Loss of Heterozygosity Tables for a detailed listing of LOH by chromosome region.)

- The fifth column lists the GenBank accession number of the target gene.
- 5 (Some of the genes specified in the Table do not yet have GenBank accession numbers. For example, genes encoding several human tRNA synthetases and ribosomal subunits have not yet been cloned, although their existence can be inferred from genetic and biochemical studies and from phylogeny.

Fig. 6 is identical to Fig. 5, except that it concerns exemplary conditionally  
 10 essential genes rather than generally essential genes.

Fig. 7 is a Target Variances Table shows molecular details of exemplary variances identified by Variagenics in exemplary target genes. There are six columns in the Table.

- The first column gives the Variagenics gene ID number, which serves as a  
 15 cross reference to the Target Variances by Field Table (see above), where information on gene location and GenBank accession number are provided. After the ID number is a decimal point and then a list of one or more integers (on successive lines), which are the (arbitrary) numbers of the specific variances identified. Between one and 13 variances were identified  
 20 per target gene. Information on different target genes is separated by dashed horizontal lines.
- The second column lists the location of the variance - specifically the number of the nucleotide at which variation was observed. The nucleotide number refers to a cDNA sequence of the target gene which can be retrieved using  
 25 the GenBank accession number provided in the Target Variances by Field Table.
- The third column lists the two variant sequences identified at the specified

- nucleotide. The variant nucleotides are bracketed and in bold font separated by a slash. Ten nucleotides of flanking sequence are provided on either side of the variance to localize the variant site unambiguously. (In the event of a conflict between the nucleotide number specified in column 2 and the sequence specified in column 3 the latter would rule as the correct sequence.)
- 5 These variances were detected by a variety of experimental and informatics based procedures described in the examples. Many variances were detected by two independent methods (e.g. informatics based detection and T4 endonuclease VII detection).
- 10 • The fourth and fifth columns (headed '# Varia 1' and '# Varia 2') provide the number of occurrences of variance 1 and 2, respectively, where variance 1 is the first and variance 2 the second of the bracketed nucleotides in column three. In both the fourth and fifth columns there are two numbers. The first number reports the number of occurrences of the variance.
- 15 'Occurrences' include ESTs identified during informatics based analysis, or variances identified experimentally by analysis of human cell lines, or both. The second number, inside parentheses, reports the number of individuals in whom the occurrences were detected. An 'individual' means either a cell line (analyzed experimentally) or a cDNA library created from one individual
- 20 (but from which many ESTs for the target gene may have been sequenced). Thus if the first number is 15 and the second number is 11 then there were 15 occurrences of the variance (a combination of 15 ESTs and/or 15 experimentally identified alleles) in a total of 11 cDNA libraries and/or cell lines.
- 25 • The fifth column provides annotation on the variances, particularly concerning the location of the variant site in the cDNA and the effect of the DNA sequence variance on the predicted amino acid sequence, if any. 5' UT = 5' untranslated region; 3' UT = 3' untranslated region; silent = variance lies in coding region by does not affect predicted amino acid

sequence; ND = analysis not done; Thr -> Asn = specific amino acid substitutions, inferred from the nucleotide sequence variance, are provided. Similar information can be readily obtained for additional genes using the methods described or as known to those skilled in the art.

5    Figures 9-15 correlate with Example 31.

**Fig. 9** is a bar graph showing the number of T24 human bladder cancer cells surviving 72 hours after transfection with antisense oligonucleotides. Anti-ras is an oligonucleotide known to have antiproliferative effects against T24 cells. This oligonucleotide exhibits inhibition comparable to the anti-RPA70 oligonucleotide.

- 10    Anti-herpes and an oligonucleotide with a scrambled sequence are shown as controls. This experiment demonstrates that RPA70 is an essential protein.

Cells were plated in six well dishes 24 hr prior to the experiment and transfected at approximately 50-70% confluency with various phosphorothioate oligomers at 400 nM. An oligomer:lipofectin ratio of 3 ug Lipofectin/ml Optimem/100 nM

- 15    Phosphorothioate oligomer was used for all transfections. Prior to transfection the cells were washed once with room temp Optimem (BRL) and then Lipofectin diluted into Optimem was added to the cells. After addition of the lipofectin the antisense oligomers were immediately added. After a five hour incubation the medium was removed from the cells and replete medium added. The cells were
- 20    allowed to recover, trypsinized, and cell number was determined at 72 hr by counting with a hemocytometer. Each bar represents two different determinations of cell number for each of three triplicate samples.

**Fig. 10** is a Northern Blot demonstrating specific suppression of RPA70 mRNA levels in two cell lines with opposite genotypes. RPA70 in Mia Paca II cells

- 25    matches the 13085 oligomer while RPA70 in T24 cells matches the 12781

oligomer. The 13706 oligomer is a random sequence control. Cells were plated in P100 dishes transfected as described in figure legend 11. Twenty-four hours after the addition of the indicated oligomers, RNA was recovered from the cells by the SDS-Lysis method (Peppel, K and Baglioni, C. *Biotechniques*, Vol. 9, No. 6, pp 711-7131, 1990). For Northern Blots 5-10 ug RNA per well was loaded onto a formaldehyde gel, electrophoresed and transferred to BioRad Zeta Probe GT. After baking (30 min at 80 C in a vac oven) the blot was probed for specific mRNA using a random primed 32P-labeled cDNA specific for RPA 70.

Fig. 11 is a Northern blot showing allele-specific Suppression of RPA 70 mRNA in T24 and Mia Paca II cells. Cells were plated in P100 dishes, transfected, and RPA 70 mRNA levels measured as previously described. T24 cells contain the genotype targeted by oligomer 12781. Mia Paca II cells are homozygous for the variance targeted by oligomer 13085. 12781 is a 20 nucleotide long phosphorothioate oligomer which targets RPA70 in T24 cells. 13085 is an 18 nucleotide long phosphorothioate oligomer which targets RPA70 in Mia Paca II cells. The lower half of the figure shows the EtBr stained gel of total RNA probed by Northern Blot.

Fig. 12 is two graphs showing that the proliferation of two cell lines homozygous for different variant forms of the RPA70 gene is inhibited to a greater degree by matched oligonucleotides than by oligomers having a single base mismatch. Cell proliferation was measured by BrdU incorporation in cellular DNA. Transfections were performed on consecutive days and BrdU incorporation measured 24 hours after the last transfection (see figure legend 9). Oligomer 12781 targets the variance contained in A549 cells and is mismatched relative to the genotype of Mia Paca II cells. Oligomer 13085 targets the variance contained in Mia Paca II cells and is mismatched relative to the genotype of A549 cells.



Fig. 13 is a graph showing Inhibition of BrdU incorporation in A549 cells by antisense oligonucleotides against the RPA 70 gene. Cells were transfected, as described previously, with a matched oligonucleotide (12781) or an oligonucleotide with one mismatch (13085). The oligonucleotide concentration was 400 nM with specific oligomer diluted with a random oligonucleotide. Cell proliferation was measured by BrdU incorporation after two transfections. Twenty-four hours after the first transfection the cells were transfected identically. Twelve hours after the second transfection BrdU was added to the cells and BrdU incorporation was assayed after a 12 hour incubation. BrdU incorporation was measured by ELISA (Boehringer Mannheim) with the following changes: Volumes were increased to assay BrdU incorporation in 6 well dishes. 1000  $\mu$ l of fix, 750 ul of antibody, and 1000 ul of substrate. A portion of the samples were transferred to a 96 well dish (in triplicate) and read at 405 nm on a plate reader.

Fig. 14 is a graph showing antiproliferative/cytopathic effects of antisense oligonucleotides against the RPA70 gene in A549 cells. Cells were transfected on three consecutive days with a matched oligonucleotide (12781) or an oligonucleotide containing a one base mismatch (13085). Following the last transfection the cells were allowed to recover three days. Cell number was quantified by Sulforhodamine B staining (Molecular Probes). Volumes were increased to accommodate the assay in 6 well dishes. Fixation 1.25 ml, stain 750 ul, solubilizer 1 ml. A portion of the samples were then transferred to a 96 well dish in triplicate and quantified by plate reader at 565 nm. All transfections were done with 400 nM oligomer by dilution of the specific oligomer with a random oligonucleotide to control for nonspecific oligonucleotide effects.

Fig. 15 is a graph showing antiproliferative/cytopathic effects in Mia Paca II cells by antisense oligonucleotides against the RPA70 gene. Cells were transfected with a matched oligonucleotide (13085) or an oligomer with a one base mismatch

(12781). Methods were identical to those described in figure legend 16.

Fig. 16 is a Northern blot showing suppression of Ribonucleotide Reductase (RR) mRNA by antisense oligomers. Mia Paca II cells were transfected and 24 hours later RR mRNA was measured by Northern Blot (for methods see figure legend 11). All oligomers have a phosphorothioate backbone throughout and are without modification. The lower half of each panel is a EtBr stained gel of the total RNA probed. Oligomer 13704 is a scrambled random control oligomer. RR2410GA targets the variance contained in Mia Paca II cells. Oligomer RR2410AG has two mismatches compared to the genotype of Mia Paca II cells. Oligomers RR1030 and RR1031 are negative control oligomers. They are targeted to a region of RR which is not effective for mRNA down-regulation.

Fig. 17 shows a Northern blot which is performed similarly to the experiments in Fig. 16. MDA-MB 468 cells were transfected and the level of RR mRNA measured after 24 hours. 13706 is a scrambled random control oligomer. 2410AG targets the two variances contained in the MDA-MB 468 cells. Oligomer 2410GA contains two mismatches relative to the genotype of MDA-MB 468 cells. Both 2410AG and 2410GA are identical to RR2410AG and RR2410GA, respectively.

Fig. 18 shows specific suppression of EPRS mRNA using hybrid oligomers. The sequences at the top provide the structures of the oligonucleotides. The graph at the bottom shows the relative specificity of oligonucleotides.

Fig. 19 is two blots showing specific suppression of EPRS mRNA using hybrid oligomers. A549 cells were transfected with the indicated concentrations of the hybrid oligomers (for structure see text). 14977 targets the two variances contained in A549 cells. 14971 contains two mismatches relative to the genotype

of A549 cells.

Fig. 20 is a graph showing inhibition of mutant *ras* using antisense oligonucleotides specific for the mutant form, based on information available in Schwab et al., 1994, PNAS 91:10460-10464.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

### I. Introduction

All normal human cells have two copies of each autosomal chromosome (chromosomes 1 through 22); one copy is inherited from each parent. Each  
5 chromosome pair thus contains two alleles for any gene. If a single allele of any gene pair is defective or absent, the surviving allele will continue to produce the encoded gene product. Generally, one allele of a gene pair is sufficient to carry on the normal functions of the cell. (Dominant genetic disorders in which mutations in one allele are sufficient to cause disease are generally those in which the mutation,  
10 or gene product harboring the mutation, has a toxic effect on the cell.)

Because humans are genetically heterogeneous, many of the paired alleles of genes of the somatic cells of an individual differ from one another in their gene sequence. Typically both alleles are transcribed and ultimately translated into proteins used by the cell. In most cases, the sequence differences between two allelic forms of a gene  
15 in an individual are small, usually differing by only one or a few base differences in sequence. The sequence differences may occur at a single variance site, or may constitute more than one variance site, *i.e.*, two allelic forms in an individual may have more than one sequence variance distinguishing them.

When a cell is heterozygous, *i.e.*, has at least one sequence variance, within the  
20 transcribed sequence for a particular gene, each allele may encode a different mRNA, *i.e.*, the mRNAs differ in base sequence. For base changes which are located within coding sequences, the effect of the nucleotide difference depends on whether the base change changes the amino acid which is encoded by the relevant codon. Many base changes do not change the coding sequence because they lie in  
25 untranslated regions of the mRNA, outside of the mRNA in introns or intergenic sequences, or in a "wobble" position of a codon which changes the codon, but not

the amino acid it encodes. As a result, the mRNAs encoded by two alleles may translate into the same protein or into forms of the same protein differing by one or more amino acids. An important aspect of the present invention is that many sequence variances that are targets for cancer therapy by the methods described here  
5 are not mutations, are not functionally related to cancer, and may not, under normal environmental conditions, induce any function difference between the allelic forms of the gene or protein. Only in the circumstances described in this invention, namely genes that encode essential functions, the presence of variances with a sufficient population frequency, a sufficient frequency of LOH in cancers, do these genes, and  
10 the variant sequences within these genes, have utility for the therapy of cancer and other disorders through the discovery of variance-specific inhibitors.

Gene targets for a variance-specific inhibition strategy in this invention satisfy three criteria:

1. The target gene encodes a gene product, *e.g.*, a RNA transcript or protein product  
15 essential for the growth or survival of cells.
  2. The target gene is located within a chromosome region frequently deleted in cancer cells or cells of a noncancer, proliferative disorder.
  3. The target gene exists in two alternative forms in the normal somatic cells of a patient having a cancer or noncancer proliferative disorder.
- 20 The allele specific therapy strategy for cancer and noncancer proliferative disorders utilizes the genetic differences between normal cells and neoplastic cells. Thus, the first step in the therapeutic strategy is identifying genes which code for proteins or other factors essential to cell survival and growth that are lost through LOH in tumor cells. Since many genes have been mapped to specific chromosomal regions, this  
25 identification can be readily performed by identifying such essential genes which are located in the chromosomal regions characteristically or frequently deleted in

different forms of human cancer or other tumors. Table 2, from the review conducted by Lasko *et al.*, 1991, *Ann. Rev. Genetics* 25:281-314, summarizes results of numerous studies determining loss of heterozygosity in tumors, identifying specific tumor types. A much larger summary of tumor-related LOH is provided in  
5 Fig. 5.

Once regions of LOH are identified in the chromosomes of a patient's tumor cells, genes which map to the deleted chromosomal segments and are known to code for gene products essential for cell growth or survival are tested for DNA sequence variances. The identification of a greater number of LOH sites affords a broader  
10 selection of target genes coding for essential proteins or other gene products and therefore of sequence variance sites for targeting.

Essential genes which have sequence variants in a population provide a set of target which are advantageous due to the presence of many patients heterozygous for a particular gene, so that the gene will provide a target in cases where the gene has  
15 undergone tumor-related LOH.

In accord with the description of target gene categories above, most advantageously a target gene is an essential gene which undergoes LOH in a tumor at a high frequency as described above and which has alternative allelic forms in a population at frequencies as described above. Such genes will provide many potentially  
20 treatable patients due to the conjunction of LOH and heterozygosity frequencies.

The most preferred target genes are those essential genes which have both a preferable rate of heterozygosity and a preferable frequency of LOH in a tumor or other proliferative condition in a population of interest. Also preferable is that the gene undergoes LOH in a plurality of different tumors or other conditions.

## II. Essential Cellular Function and Essential Genes

As indicated in the Summary above, the invention targets specific allelic forms of essential genes, which are also termed genes essential for cell growth or viability.

As used herein the term, "genes which code for a protein essential for the growth or survival of cells" or "genes which code for proteins or factors required for cell viability" or "essential genes" is meant to include those genes that express gene products (*e.g.*, proteins) required for cell survival as well as those genes required for cell growth in actively dividing cell populations. These genes encode proteins which can be involved in any vital cell. An additional factor which applies to genes identified by any of the approaches described above is: a target gene or protein should be encoded by a single locus in man.

A large number of references have identified essential genes which constitute actual or potential targets for allele specific inhibition. The identification of essential genes can be approached in various ways.

1. What are the essential functions each cell must perform to sustain life, and what are the proteins responsible for performing those functions? This is a top down approach for identifying candidate genes whose essential role is then proven experimentally (see below). This approach enables essential genes to be categorized according to the essential cellular process or function which the gene product provides or of which the gene product is a necessary part. Table 1 shows such categories of essential genes and gene functions. In addition, the chromosomal location, where known, and gene product of certain example genes is provided. Thus, the categories of functions shown provide potential targets for the methods of this invention.
2. What genes have been proven essential for cell survival by mutagenesis or gene disruption experiments in cells of other organisms, such as hamster cells, mice,

- flies, yeast, bacteria or other organisms? The idea of determining the necessity of specific genes for survival of an organism is well established in simple organisms such as bacteria and yeast. The consequences of gene disruption are easier to assess in these microorganisms that have a haploid genome because the
- 5 haploid organism contains only one form of a particular single copy gene. A particularly useful category of eukaryotic organisms are the yeasts, especially *Saccharomyces cerevisiae*.
3. What are the protein targets of proven mammalian cytostatic and cytotoxic agents such as chemotherapy drugs and poisons?
- 10 4. What can be learned from genomics about the genes required for cell survival? This analysis includes identification of the minimal gene set in simple prokaryotes, as well as sequence comparisons across widely divergent species.
5. Experimental testing of gene essentiality. As an example, antisense oligonucleotides can be used to down regulate candidate essential genes
- 15 (identified by the four approaches listed above) and assess the effects on cell proliferation and survival. Application of an antisense approach to the identification of essential genes was described by Pestov & Lau, *supra*.

Once a gene coding for a protein or factor essential to cell viability is identified, its genomic DNA and cDNA sequences, if not previously established, can be

20 ascertained and sequenced according to standard techniques known to those skilled in the art. See, for example, Sambrook, Fritsch and Maniatis, "Molecular Cloning, A Laboratory Manual," Cold Spring Harbor Press, Cold Spring Harbor, NY (1989).

#### *Categories of essential genes*

Many essential genes function by encoding a gene product which is necessary for

25 maintaining the level of a cellular constituent within the levels required for cell survival or proliferation. The survival and proliferation of cells within the body requires maintaining a state of homeostasis among many different cellular



constituents. These may include, but are not limited to, specific proteins, nucleic acids, carbohydrates, lipids, organic ions, and inorganic ions, or cytoskeletal elements. The loss of homeostasis often results in cell death or apoptosis or inhibition of cell proliferation. Homeostasis in a living cell is dynamic, and

5   programed changes in homeostasis are required through the life cycle of the cell. We have determined that those genes whose products are required for maintaining this homeostasis conducive to cell growth and survival are targets for anti-neoplastic *e.g.*, anti-cancer, inhibitors as described in the methods herein. For example, many genes are involved in synthetic functions, allowing the cells to produce essential

10   cellular constituents including proteins, nucleic acids, carbohydrates, lipids, or organic ions or their components. Other genes are involved in the transport of essential constituents such as proteins, nucleic acids, carbohydrates, lipids, organic ions, or inorganic ions, or their components into the cell or among its internal compartments. Still other genes are involved in the chemical modification of

15   cellular constituents to form other constituents with specific activities. Still other genes are involved in the elimination of specific cellular constituents such as proteins, nucleic acids, carbohydrates, lipids, organic ions, inorganic ions, or their components by metabolic degradation or transport out of the cell. The analysis is preferably carried out using genes which have been shown to be essential in human

20   cells or which are human homologs of genes which are essential in other organisms, preferably other eukaryotic organisms although useful essential data is also provided by prokaryotic essential genes.

A specific example are those genes that are involved in maintaining the amount and fidelity of DNA within a cell. This includes genes commonly considered to be

25   involved in "replication" and other functions; comprising genes involved in the synthesis (polymerization) of DNA sequences from its component elements, creating specific modifications of DNA, ensuring the proper compartmentalization of DNA during cell division (within the nucleus), and eliminating damaged DNA.

This also includes those genes involved in maintaining the amount of nucleosides that are the component elements of DNA by synthesis, salvage, or transport.

Another example are those genes that are involved in maintaining the amount of RNAs within a cell. This includes genes commonly considered to be involved in transcription and other functions; comprising genes required for the synthesis (polymerization) of linear RNA sequences from its component elements, ensuring proper compartmentalization of RNA within the cell, creating specific modification of the linear RNA molecule, and eliminating RNA. This also includes those genes involved in maintaining the amount of nucleosides that are the component elements of RNA by synthesis, salvage, or transport.

Another example are those genes that are involved in maintaining the amount of proteins within a cell. This includes those genes commonly considered to be part of "translation" and other functions; comprising genes required for transporting or synthesizing amino acids that are the component elements of proteins, synthesizing specific linear protein sequences from these amino acid elements, creating specific modifications of proteins including by not limited to the addition of specific nucleic acids, carbohydrates, lipids, or inorganic ions to the protein structure, ensuring the proper compartmentalization of synthesized proteins in the cell, and ensuring the proper elimination of proteins from the cell.

Another example are those genes that are involved in maintaining the amount of organic ions within the cell, including but not limited to amino acids, organic acids, fatty acids, nucleosides, and vitamins. This includes those genes that are required for transporting, or synthesizing organic ions, ensuring their proper compartmentalization within the cell, and ensuring proper elimination or degradation of these ions.

Another example are those genes that are involved in maintaining the amount of inorganic ions within the cell. This includes those genes that are required for transporting inorganic ions, including but not limited to O, Na, K, Cl, Fe, P, S, Mn, Mg, Ca, H, PO<sub>4</sub> and Zn, ensuring their proper compartmentalization within the cell  
5 by binding or transporting these ions, and ensuring proper elimination from the cell.

Another example are those genes that are involved in maintaining the structures and integrity of the cell as described in Example 6 below.

The above groups of genes are shown in Table 1 below, which also points out useful subcategories of genes and lists particular exemplary target genes. This  
10 demonstrates that target genes can be grouped according to cellular function to provide classes of essential genes useful for allele specific targeting. Additional target genes can be identified by routing methods, such as those described herein. Confirmation of the essentiality of an additional gene in a specified gene category, and of the occurrence in normal somatic cells of sequence variances of the gene, and  
15 of the occurrence of LOH affecting the gene in a neoplastic disorder, establishes that the gene is a target gene potentially useful for identifying allele specific inhibitors and for other aspects of the invention. In addition, as described, target genes are useful in embodiments of certain aspects of the invention, e.g., transplantation and the use of essential or conditionally essential genes even in the absence of LOH.

20

**Table 1**

Gene Name	GenBank Accession #
1) Genes Required For Cell Proliferation	

## 1.1 Genes that regulate cell division

Cyclins, cyclin dependent kinases, regulators and effectors of cyclins and cyclin-dependent kinases

14-3-3 Protein TAU	X56468
CCNA(G2/Mitotic-Specific Cyclin A )	X51688
CCNB1(G2/Mitotic-Specific Cyclin B1)	M25753
CCND1(G1/S-Specific Cyclin D1)	M73554
CCND2(G1/S-Specific Cyclin D2)	M90813
CCND3(G1/S-Specific Cyclin D3)	M90814
Cell division control protein 16	U18291
Cell division cycle 2, G1 to S and G2 to M	X05360
Cell division cycle 25A	M81933
Cell division cycle 25B	M81935
Cell division cycle 25C	M34065
Cell division cycle 27	U00001
Cell division-associated protein BIMB	D79987
Cyclin A1(G2/Mitotic-Specific Cyclin A1 )	U66838
Cyclin C (G1/S-Specific Cyclin C)	M74091
Cyclin G1(G2/Mitotic-Specific Cyclin G)	X77794
Cyclin G2 (G2/Mitotic-Specific Cyclin G)	U47414
Cyclin H	U11791
Cyclin H Assembly	X87843
GSPT1(G1 to S phase transition 1)	X17644
Mitotic MAD2 Protein	U31278
MRNP7	X98263
RANBP1(RAN binding protein 1)	D38076
WEE1	X62048
Cell Division Protein Kinase 4	U79269
CDC28 protein kinase 1	X54941
CDC28 protein kinase 2	X54942
M-Phase inducer phosphatase 2	M81934
M-phase phosphoprotein, mpp6	X98260
PPP1ca(Protein phosphatase 1, catalytic subunit, alpha isoform)	M63960
STM7-LSB	X92493

1.2 Genes that form structures of cell division including the centromere, kinetochore, kinesins, spindle pole body, chromatin assembly factors and their regulators

CENP-F kinetochore protein	U19769
Centromere autoantigen C	M95724

74

232/116

Centromere protein B (80kD)	X05299
Centromere protein E (312kD)	Z15005
CHC1(Chromosome condensation 1)	X12654
Chromatin assembly factor-in p150 subunit	U20979
Chromatin assembly factor-in p60 subunit	U20980
Chromosome segregation gene homolog CAS	U33286
HMG1(High-mobility group (nonhistone chromosomal) protein 1)	D63874
Minichromosome Maintenance (MCM7)	D28480
Mitotic centromere-associated kinesin	U63743
RMSA1(Regulator of mitotic spindle assembly 1)	L26953
SUPT5h(Chromatin structural protein homolog (SUPT5H))	Y12790

## 2) Genes Required to Maintain Inorganic Ions and Vitamins at Levels Compatible with Cell Growth or Survival

### 2.1 Transport of inorganic ions and vitamins across the plasma membrane and intracellular membranes

#### Active transporters

##### Uniporters

PMCA1 (Calcium Pump)	U15686
PMCA2 (Calcium Pump)	M97260
PMCA3 (Calcium Pump)	U15689
PMCA4 (Calcium Pump)	M83363
ATP2b1 (Calcium-Transporting ATPase Plasma Membrane)	J04027
ATP2b2 (Calcium-Transporting ATPase Plasma Membrane)	X63575
ATP2b4 (Calcium-Transporting ATPase Plasma Membrane)	M83363
ATP5b (ATP Synthase Beta Chain, Mitochondrial Precursor )	X03559
Chloride Conductance Regulatory Protein ICLN	X91788
H-Erg (Potassium Channel Protein EAG)	U04270
Nuclear Chloride Ion Channel Protein (NCC27)	U93205
SCN1b(Sodium Channel, Voltage-Gated, Type in, Beta Polypeptide)	L16242
Two P-Domain K <sup>+</sup> Channel TWIK-1	U33632
VDAC2 (Voltage-Dependent Anion-Selective Channel Protein 2)	L06328

#### Coupled transporters

##### Symporters

ATP1b1 (Sodium/Potassium-Transporting ATPase Beta-1 Chain)	X03747
--	--------

75

232/116

ATP1b2 (Sodium/Potassium-Transporting ATPase Beta-2 Chain)	M81181
Antiporters	
ATPase, Ca <sup>++</sup> transporting, plasma membrane 4	M25874
ATPase, Ca <sup>++</sup> transporting, plasma membrane 2	L20977
ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, alpha 1 polypeptide	U16798
ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, alpha 3 polypeptide	X12910
ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, beta 1 polypeptide	U16799
ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, beta 2 polypeptide	U45945
Na <sup>+</sup> ,K <sup>+</sup> ATPase, 1 Subunit	
Na <sup>+</sup> ,K <sup>+</sup> ATPase, 2 alpha	
Na <sup>+</sup> ,K <sup>+</sup> ATPase, 3 beta	U51478
SLC9a1(Solute carrier family 9 (sodium/hydrogen exchanger))	M81768
Solute carrier family 4, anion exchanger, member 1	M27819
Solute carrier family 4, anion exchanger, member 2	U62531
Solute carrier family 9 (sodium/hydrogen exchanger),	X76180
Passive transporters	
MaxiK Potassium Channel Beta Subunit	U25138
Chloride Channel 2	X83378
Chloride Channel Protein (CLCN7)	U88844
TRPC1 (Transient Receptor Potential Channel 1)	X89066
Potassium Channel Kv2.1	L02840
ATP5d(ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, delta subunit)	X63422
ATP5f1(ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit b)	X60221
ATP5o(ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, O subunit)	X83218
ETFa(Electron-transfer-flavoprotein, alpha polypeptide (glutaric aciduria II))	J04058
ETFb(Electron-transfer-flavoprotein, beta polypeptide)	X71129
Nadh-ubiquinone oxidoreductase 13 kd-B subunit	U53468
Nadh-ubiquinone oxidoreductase 39 kD subunit precursor	L04490

76

232/116

NADH-Ubiquinone oxidoreductase 75 kD subunit precursor	X61100
NADH-Ubiquinone oxidoreductase MFWE subunit	X81900
NDUFV2(NADH dehydrogenase (ubiquinone) flavoprotein 2 (24kD))	M22538
Ubiquinol-cytochrome c reductase complex 11 kD	M36647
ATP Synthase Alpha Chain	D14710
NADH dehydrogenase-ubiquinone Fe-S protein 8, 23 kDa subunit	U65579

## Vitamin transporters

Ascorbic Acid (uncloned)	
Folate Binding Protein	AF000380
Folate receptor 1 (adult)	M28099
Nicotinamide (uncloned)	
Pantothenic Acid	X92762
Riboflavin (uncloned)	
SCL19A1 (Solute Carrier Family 19, Member1)	
Solute carrier family 19 (folate transporter), member 1	U19720
Thiamine, B6, B12 (uncloned)	

## Metal transporters

ATP7b (Copper-Transporting ATPase 2)	U03464
Ceruloplasmin (ferroxidase)	M13699
Ceruloplasmin receptor (Copper Transporter)	
Copper Transport Protein HAH1	U70660
Molybdenum, Selenium, other Transporters (uncloned)	
Transferrin Receptor (Iron Transporter)	X01060
Zinc Transporter (uncloned)	

## Soluble inorganic ion transporters

## Insoluble inorganic ion transporters

## Transporters of other essential small molecules

Mitochondrial Import Receptor Subunit TOM20	D13641
--	--------

## 2.2 Regulators of transport

## Sensors of ion levels

### 3) Genes Required to Maintain Organic Compounds at Levels Compatible with Cell Growth or Survival

## 3.1 Transporters of organic compounds

## Carbohydrate Transport

## Sugar Transport

## Glucose Transport

77

232/116

GLUT1	GDB:120627
GLUT2	J03810
GLUT3	M20681
GLUT4	M20747
GLUT5	M55531
GLUT6	M95549
Solute carrier family 5 (sodium/glucose cotransporter)	M95549
Solute carrier family 2 (facilitated glucose transporter), member 2	J03810
Solute carrier family 2 (facilitated glucose transporter) member 5	M55531
Amino acid transport	
Solute carrier family 3 member 1	L11696
System b,(Na <sup>+</sup> independent)	
System y,(Na <sup>+</sup> independent)	
ATRC1(Catioinc)	OMIM 104615
LEUT(Leucine Transporter)	OMIM 151310
SLC1A1(Solute Carrier Family 1, Member 1)	OMIM 133550
Lipid or lipoprotein transport	
Nucleoside transport	
Other organic compounds transport	
Solute carrier family 16 (monocarboxylic acid transporters)	L31801
3.2 Genes required for maintenance of organic compounds at levels required for cell growth or survival	
Carbohydrate metabolism, including anabolism and catabolism	
ACO1(Aconitase 1)	
ACO2(Aconitase 2, mitochondrial)	U80040
Acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	M26393
Acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	M16827
Acyl-Coenzyme A dehydrogenase, long chain	M74096
Acyl-Coenzyme A dehydrogenase, very long chain	D43682
aKGD (alpha ketoglutaratedehydrogenase)	
ALD-a (Aldolase)	M11560
ALD-b (Aldolase)	K01177
ALD-c (Aldolase)	M21191
CS (Citrate Synthetase)	OMIM 118950
Dihydrolipoamide S-succinyltransferase	L37418
DLAT(Dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex))	AF001437
DLD(Dihydrolipoamide dehydrogenase (E3 component of pyruvate dehydrogenase complex, 2-oxo-glutarate complex, branched chain keto acid dehydrogenase complex))	J03490
Elk (Oxoglutarate dehydrogenase)	D10523



E2k (Dihydrolipoamide S-succinyltransferase)	D16373
E3 (Dihydrolipoyl Dehydrogenase)	SEG_HUMDHL
ENO1(Enolase 1, alpha)	M14328
ENO2(Enolase 2)	X51956
ENO3(Enolase 3)	X55976
Enolase 2, (gamma, neuronal)	M22349
Enolase 3, (beta, muscle)	X16504
FH(Fumarate hydratase)	M15502
G3PDH (Glyceraldehyde-3-Phosphate Dehydrogenase)	M17851
G6PD (Glucose-6-Phosphate Dehydrogenase)	
Glucose-6-phosphate dehydrogenase	X03674
HK1 (Hexokinase 1)	M75126
HK2 (Hexokinase 2)	S70035
HK3 (Hexokinase 3)	U51333
IDH1(Isocitrate dehydrogenase 1 (NADP+), soluble)	OMIM 147700
IDH2(Isocitrate dehydrogenase 2 (NADP+), mitochondrial)	X69433
MDH1(Malate dehydrogenase 1, NAD (soluble))	D55654
MDH2(Malate dehydrogenase 1, NAD (mitochondrial))	OMIM 154100
NAD(H)-specific isocitrate dehydrogenase alpha subunit	U07681
Oxoglutarate dehydrogenase (lipoamide)	D10523
PDHB (Pyruvate Dehydrogenase)	J03576
PDHB(Pyruvate dehydrogenase (lipoamide) beta)	M34479
PDK4 (Pyruvate dehydrogenase kinase, isoenzyme 4)	U54617
PFKL(Phosphofructokinase)	M10036
PGI (Phosphoglucosomerase)	OMIM 172400
PGKa (Phosphoglyceromutase)	Y00572
PGKb (Phosphoglyceromutase)	K03201
PGM1 (Phosphoglyceromutase)	M83088
PGM2 (Phosphoglyceromutase)	OMIM 172000
PGM3 (Phosphoglyceromutase)	OMIM 172100
PGM4 (Phosphoglyceromutase)	OMIM 172110
Phosphofructokinase, muscle	U24183
Phosphoglucomutase 1	M83088
Phosphoglycerate kinase 1	V00572
PK1 (Pyruvate Kinase)	M15465
PK2 (Pyruvate Kinase)	OMIM 179040
PK3 (Pyruvate Kinase)	M23725
Pyruvate dehydrogenase kinase isoenzyme 2 (PDK2)	L42451
Pyruvate kinase, liver	D10326
Pyruvate kinase, muscle	M23725
SDH1(Succinate dehydrogenase, iron sulphur (Ip) subunit)	D10245
SDH2(Succinate dehydrogenase 2, flavoprotein (Fp) subunit)	D30648
TKT(Transketolase (Wernicke-Korsakoff syndrome))	L12711
TP1 (Triseposphate Isomerase)	M10036

79

232/116

Amino Acid biosynthesis and processing	
Asparagine Synthetase	SEG_HUMASN
Aminoacylase-1	L07548
Aminoacylase-2	S67156
Fatty acid biosynthesis and processing	
ACAC (Acetyl CoA Carboxylase Beta)	U19822
ACAC (Acetyl CoA Carboxylase)	U12778
ACADSB(Acyl-coA dehydrogenase)	U12778
Mevalonate kinase	M88468
Phosphomevalonate kinase	L77213
Alcohol biosynthesis and processing	
Other organic compounds biosynthesis and processing	
Aspartoacylase	S67156
Ornithine decarboxylase 1	M16650
3.3 Genes required for catabolism, degradation and elimination of organic compounds	
Carbohydrate and Sugar Catabolism	
Amino acid Degradation	
Lipid or lipoprotein Degradation	
Short-acyl-CoA dehydrogenase	M26393
Medium acyl-CoA dehydrogenase	S75214
Long acyl-CoA dehydrogenase	M74096
Isovaleryl CoA dehydrogenase	M34192
2-methyl branched chain	
Nucleoside Degradation	
Adenosine Deaminase	K00509
Purine-nucleoside phosphorylase	K02574
Guanine Deaminase	
Xanthine Oxidase	D11456
Degradation of other organic compounds	
3.4 Genes Required to Modify Polypeptides, Lipids or Sugars by Addition, Removal or Modification of Chemical Groups to Form Compounds Necessary for Cell Growth or Survival	
Addition, removal or modification of sugar groups	
Glycosyltransferases	
Glycosylases	
ITM1 (Integral Transmembrane Protein)	L38961
GFPT (Glutamine-Fructose-6-Phosphate Transaminase)	M90516
Heparan	U36601
Polypeptide N-Acetyltransferase	U41514
Addition, removal or modification of methyl or other alkylgroups	
Acetyltransferase	
ACAA(Acetyl-Coenzyme A acyltransferase)	X12966
Lysophosphatidic acid acyltransferase-alpha	U56417

80

232/116

Lysophosphatidic acid acyltransferase-beta	U56418
Farnesyltransferase	
FNTa (Farnesyltransferase Alpha Subunit)	L00634
FNTb (Farnesyltransferase Beta Subunit)	L00635
Myristoylation	
NMT1 (N-myristoyltransferase)	
Addition, removal or modification of sulfhydryl groups	
Addition, removal or modification of phosphate groups	
Calcineurin A	S46622
Calcineurin B	M30773
Calreticulin Precursor	M84739
Phosphatase 2b	M29551
PPP3ca(Protein phosphatase 3 , catalytic subunit)	J05480
SNK Interacting 2-28(Calcineurin B Subunit)	U83236
Protein Kinase C	
PRKCA(Protein kinase C, alpha)	X52479
PRKCB1(Protein kinase C, beta 1)	X06318
PRKCD(Protein kinase C, delta)	L07861
PRKCM(Protein kinase C, mu)	X75756
PRKCQ(Protein kinase C-theta)	L01087
PRKCSH(Protein kinase C substrate 80K-H)	J03075
Addition, removal or modification of lipid groups	
Geranylgeranyl	
Geranylgeranyltransferase (Type I Beta)	L25441
GGTB (Geranylgeranyltransferase)	Y08201
Geranylgeranyltransferase (Type II Beta-Subunit)	X98001
3.5 Genes required for regulation of levels of organic ions	
Gdp Dissociation Inhibitors	
GDI Alpha (RAB GDP Dissociation Inhibitor Alpha)	D45021
Rab Gdp (RAB GDP Dissociation Inhibitor Alpha)	D13988
<b>4) Genes Required to Maintain Cellular Proteins at Levels Compatible with Cell Growth or Survival</b>	
Polypeptide precursor biosynthesis	
Amino acid biosynthesis and modification	
GOT(Glutamic-oxaloacetic transaminase 2)	M22632
GOT1(Glutamic-oxaloacetic transaminase 1)	M37400
PYCS(Pyrroline-5-carboxylate synthetase)	X94453
Tyrosine aminotransferase	X52520
Polypeptide precursor elimination	
Synthesis of components for polypeptide polymerization	
AARS	D32050
CARS	L06845
DARS	

81

232/116

EPRS	X54326
FARS	
GARS	U09510
HARS	X05345
IARS	D28473
KARS	OMIM 601421
LARS	OMIM 151350
MARS	X94754
NARS	M27396
QARS	X54326
RARS	S80343
SARS	
TARS	M63180
VARs	X59303
WRS	M61715
YARS	

## Polypeptide polymerization

## Ribosome Subunits

Ribosomal Protein L11	X79234
Ribosomal Protein L12	L06505
Ribosomal Protein L17	X52839
Ribosomal Protein L18	L11566
Ribosomal Protein L18a	X80822
Ribosomal Protein L19	X63527
Ribosomal Protein L21	U14967
Ribosomal Protein L22	L21756
Ribosomal Protein L23	X53777
Ribosomal Protein L23a	U43701
Ribosomal Protein L25	
Ribosomal Protein L26	
Ribosomal Protein L27	L19527
Ribosomal Protein L27a	U14968
Ribosomal Protein L28	U14969
Ribosomal Protein L29	U10248
Ribosomal Protein L30	OMIM 180467
Ribosomal Protein L31	
Ribosomal Protein L32	X03342
Ribosomal Protein L35	U12465
Ribosomal Protein L35a	X52966
Ribosomal Protein L36a	OMIM 180469
Ribosomal Protein L39	U57846
Ribosomal Protein L4	L20868
Ribosomal Protein L41	
Ribosomal Protein L44	

Ribosomal Protein L6	X69391
Ribosomal Protein L7	L16558
Ribosomal Protein L7a	X52138
Ribosomal Protein L8	Z28407
Ribosomal Protein L9	U09953
Ribosomal Protein P1	M17886
Ribosomal Protein S10	U14972
Ribosomal Protein S11	X06617
Ribosomal Protein S13	L01124
Ribosomal Protein S14	
Ribosomal Protein S15	J02984
Ribosomal Protein S15A	X84407
Ribosomal Protein S16	M60854
Ribosomal Protein S17	M13932
Ribosomal Protein S17A	OMIM 180461
Ribosomal Protein S17B	OMIM 180462
Ribosomal Protein S18	L06432
Ribosomal Protein S20	
Ribosomal Protein S20A	OMIM 180463
Ribosomal Protein S20B	OMIM 180464
Ribosomal Protein S21	L04483
Ribosomal Protein S23	D14530
Ribosomal Protein S25	M64716
Ribosomal Protein S26	X69654
Ribosomal Protein S28	U58682
Ribosomal Protein S29	L31610
Ribosomal Protein S3	U14990
Ribosomal Protein S3A	OMIM 180478
Ribosomal Protein S4	
Ribosomal Protein S4X	M58458
Ribosomal Protein S4Y	M58459
Ribosomal Protein S5	U14970
Ribosomal Protein S6	J03537
Ribosomal Protein S7	M77233
Ribosomal Protein S8	OMIM 600357
Ribosomal Protein S9	U14971
Initiation of polypeptide polymerization	
eIF-2 (Eukaryotic initiation factor)	L19161
eIF-2-associated p67(Eukaryotic initiation factor)	U13261
eIF-2A(Eukaryotic initiation factor)	J02645
eIF-2Alpha(Eukaryotic initiation factor)	U26032
eIF-2B(Eukaryotic initiation factor)	U23028
eIF-2B-Gamma(Eukaryotic initiation factor)	L40395
eIF-2Beta(Eukaryotic initiation factor)	M29536

83

232/116

eIF-3 p110(Eukaryotic initiation factor)	U46025
eIF-3 p36(Eukaryotic initiation factor)	U39067
eIF-4A(Eukaryotic initiation factor)	D21853
eIF-4C(Eukaryotic initiation factor)	L18960
eIF-4E(Eukaryotic initiation factor)	M15353
eIF-4Gamma(Eukaryotic initiation factor)	Z34918
eIF-5(Eukaryotic initiation factor)	U49436
eIF-5A	
Polypeptide elongation	
Eukaryotic peptide chain release factor subunit 1	X81625
P97(Eukaryotic initiation factor)	U73824
eEF1A2(Eukaryotic elongation factor)	X70940
eEF1D(Eukaryotic elongation factor)	Z21507
eEF2(Eukaryotic elongation factor)	X54166
eIF4A2 (Eukaryotic initiation factor)	D30655
KIAA0031(Elongation factor 2)	D21163
KIAA0219(Putative translational activator C18G6.05C)	D86973
Factor 1-Alpha 2(Eukaryotic translation elongation factor 1 alpha 2)	D30655
Termination of polypeptide polymerization	
Polypeptide folding	
Cis-Trans Isomerase	M80254
DNAj Protein Homolog 1	X62421
DNAj Protein Homolog 2	D13388
DNAJ Protein homolog HSI1	X63368
Chaperone proteins	
T-Complex	
Aspartylglucosaminidase	X55330
T-Complex 1, Alpha	S70154
T-Complex 1, Epsilon	D43950
T-Complex 1, Gamma	X74801
T-Complex 1, Theta	D13627
T-Complex 1, Zeta	M94083
Polypeptide Degradation	
Proteasome components and proteinases	
26S Protease regulatory subunit 4	L02426
Alpha-2-Macroglobulin	M11313
Calpain 1, Large	X04366
CLPP(ATP-Dependent CLP protease proteolytic subunit)	Z50853
KIAA0123 (Mitochondrial processing peptidase alpha subunit)	D50913
MMP7	X07819
Proteasome Beta 6	D29012
Proteasome Beta 7	D38048
Proteasome C13	U17496

84

232/116

Proteasome C2	D00759
Proteasome C7-1	D26599
Proteasome inhibitor hPI3I subunit	D88378
Proteasome P112	D44466
Proteasome P27	AB003177
Proteasome P55	AB003103
Ubiquitin System	
Enzyme E2-17 Kd(Cyclin-selective ubiquitin carrier protein)	U73379
ISOT-3(Ubiquitin carboxyl-terminal hydrolase T )	U75362
ORF (Ubiquitin carboxyl-terminal hydrolase 14)	M68864
PGP(Ubiquitin carboxyl-terminal hydrolase isozyme L1)	X04741
UBA52(Ubiquitin A-52 residue ribosomal protein fusion product 1)	S79522
Ubiquitin carboxyl-terminal hydrolase 3	D80012
Ubiquitin carboxyl-terminal hydrolase isozyme L3	M30496
Ubiquitin carboxyl-terminal hydrolase T	X91349
Ubiquitin carrier protein (E2-EPF)	M91670
Ubiquitin fusion-degradation protein (UFD1L)	U64444
Ubiquitin Hydrolase	X98296
Ubiquitin-conjugating enzyme E2I	U45328
Polypeptide Transport	
SEC23(Protein transport protein SEC23)	X97065
SEC23A(Protein transport protein SEC23)	X97064
SEC7(Protein transport protein SEC7)	X99688
SEC61 (Beta Subunit)	L25085
Lipoprotein Transport	
LDLR (LDL receptor)	

#### 5) Genes Required to maintain Cellular Nucleotides at Levels Compatible with Cell Growth or Survival

##### Genes Required to Maintain Cellular DNA with Fidelity and at Levels Compatible with Cell Growth or Survival

##### DNA Precursor Biosynthesis

Adenylate Kinase-2	U39945
Adenylosuccinate synthetase	X66503
Adenylosuccinate Lyase	X65867
ADPRT (ADP-Ribosyltransferase)	M32721
ADSL (Adenylosuccinate lyase/AMP synthetase)	X65867
ADSS (Adenylosuccinate Synthetase)	X66503
CAD PROTEIN	D78586
CTP Synthetase	
CTPS(CTP synthetase)	X52142
Cytidine Triphosphate Synthetase	
GARS (Phosphoribosylglycinamide synthetase)	D32051

85

232/116

GART (Phosphoribosylglycinamide formyltransferase)	
GART(Phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase)	X54199
GMP Synthetase	U10860
IMP Cyclohydrolase	U37436
IMP dehydrogenase	L19709
IMPDH1(IMP (inosine monophosphate) dehydrogenase 1)	J05272
IMPDH2(IMP (inosine monophosphate) dehydrogenase 2)	J04208
Phosphoribosyl diphosphotransferase	
Phosphoribosylaminoimidazolecarboxamide formyltransferase	
Phosphoribosylformylglycinamide synthetase	M32082
Phosphoribosylglycinamide carboxylase	
Phosphoribosylglycinamide-succinocarboxamide synthetase	
PPAT (Amidophosphoribosyltransferase)	
PPAT(Phosphoribosyl pyrophosphate amidotransferase)	U00238
Ribonucleoside-diphosphate reductase M1 chain	X59543
Ribonucleoside-diphosphate reductase M2 chain	X59618
Thymidine Kinase	K02581
Thymidylate Synthase	X02308
UMK(Uridine kinase)	D78335
UMPK (Uridine monophosphate kinase)	OMIM 191710
UMPS(Uridine monophosphate synthetase (orotate phosphoribosyl transferase and orotidine-5'-decarboxylase))	J03626
Uridine Phosphorylase	X90858
DNA Precursor Elimination	
DNA Replication	
Origin Recognition	
Origin Recognition Complex	
ORC1	U40152
ORC2	U27459
ORC3	
ORC4	
ORC5	OMIM 602331
ORC6	
ORC Regulators	
CDC6	AA830372
CDC7	AFO15592
CDC18	AF022109
DNA Polymerization	
DNA Polymerases	
Adprt (NAD(+) ADP- Ribosyltransferase)	M18112
DNA Polymerase Alpha-Subunit	X06745
DNA Polymerase Delta	U21090



86

232/116

POLa(DNA Polymerase Alpha/Primase Associated Subunit)	L24559
POLb(DNA Polymerase Beta Subunit)	D29013
POLd1(Polymerase (DNA directed), Delta 1, Catalytic Subunit)	M81735
POLd2(Polymerase (DNA directed), Delta 2)	U21090
POLE(Polymerase (DNA directed))	OMIM 174762
POLg (DNA Polymerase Gamma Subunit)	X98093
Terminal Transferase (DNA Nucleotidyltransferase )	M11722
Accessory factors for DNA Polymerization	
Activator I 36 Kd	L07540
CDC46 (DNA Replication Licensing Factor)	X74795
CDC47 (DNA Replication Licensing Factor CDC47)	D55716
DNA Topoisomerase III	U43431
DRAP1 (DNA Replication Licensing Factor MCM3)	U41843
KIAA0030 Gene (Cell Division Control Protein 19)	X67334
KIAA0083 Gene (DNA Replication Helicase DNA2 )	D42046
MCM3 (DNA Replication Licensing Factor MCM3)	D38073
PCNA (Proliferating Cell Nuclear Antigen)	J04718
PRIM1 (DNA Primase 49 kD Subunit )	X74330
PRIM2 (DNA Primase)	X74331
PRIM2a (DNA Primase 58 kD Subunit)	X74331
PRIM2b (DNA Primase)	OMIM 600741
RECa (Replication Protein A 14 kD Subunit)	L07493
RFC1 (Replication Factor C (activator I) 1)	L14922
RFC2 (Replication Factor C 2)	M87338
RFC3 (Replication Factor C (activator I) 3)	L07541
RFC4 (Replication Factor C, 37-kD subunit)	M87339
RFC5 (Replication Factor C)	OMIM 600407
RPA1 (Replication protein A1 (70kD))	M63488
RPA2 (Replication protein A2 (32kD))	J05249
RPA3 (Replication protein A3 (14kD))	L07493
TOP1 (DNA Topoisomerase I)	J03250
TOP2a (Topoisomerase (DNA) II Alpha (170kD))	J04088
TOP2b (Topoisomerase (DNA) II Beta (180kD))	U54831
DNA Helicases	
CHL1(CHLI-Related Helicase)	U33833
DNA Helicase II	M30938
Mi-2(Chromodomain-Helicase- DNA-Binding Protein CHD-1 )	X86691
RECQL (ATP-Dependent DNA Helicase Q1)	L36140
Smbp2 (DNA-Binding Protein SMUBP-2)	L14754
DNA Packaging Proteins	
Histones	
H1(0) (Histone H5A)	X03473
Histone H1d	X57129
Histone H1x	D64142

87

232/116

Histone H2a.1	U90551
Histone H2a.2	L19779
Histone H2b.1	M60756
Histone H4	X60486
SLBP (Histone Hairpin-Binding Protein )	Z71188
DNA Degradation	
DNA Repair	

Genes Required to Maintain Cellular RNA at Levels Compatible  
with Cell Growth or Survival

RNA Precursor Biosynthesis

RNA Precursor Elimination

RNA Polymerization

Initiation of polymerization

TATA-binding Complex

Small Nuclear RNA-Activating Complex, Polypeptide 1, 43KD Z47542  
(SNAPC1)

Small Nuclear RNA-Activating Complex, Polypeptide 2,  
(SNAPC2)

Small Nuclear RNA\_Activating Complex, Polypeptide 3, 50KD U71300  
(SNAPC3)

TAF2D(TBP-associated factor) U78525

TAFII100(TBP-associated factor) X95525

TAFII130(TBP-associated factor) U75308

TAFII20(TBP-associated factor) X84002

TAFII250(TBP-associated factor) D90359

TAFII28(TBP-associated factor) X83928

TAFII30(TBP-associated factor) U13991

TAFII32(TBP-associated factor) U21858

TAFII40(TBP-associated factor)

TAFII55(TBP-associated factor) U18062

TAFII80(TBP-associated factor) U31659

TBP(TATA Binding Protein) M55654

TMFI (TATA Element Modulatory Factor I)

Polymerization

RPB 7.0 U52427

RPB 7.6

RPB 17

RPB 14.4

RNA Polymerase I subunits

RNA polymerase I subunit hRPA39 AF008442

RNA Polymerase II subunits

13.6 Kd Polypeptide (DNA-Directed RNA Polymerase II 13.6  
kD Polypeptide) L37127

POLR2C(RNA polymerase II, polypeptide C (33kD)) J05448

88

232/116

Polypeptide A (220kd)	X63564
RNA Polymerase II 23k	J04965
RNA polymerase II holoenzyme component (SRB7)	U46837
RNA polymerase II subunit (hsRPB10)	U37690
RNA polymerase II subunit (hsRPB8)	U37689
RNA polymerase II subunit hsRPB4	U85510
RNA polymerase II subunit hsRPB7	U20659
RNA Polymerase II Subunit(DNA- Directed RNA Polymerases I, II, and III 7.3 kD polypeptide)	Z47727
TCEB1L(Transcription elongation factor B (SIII), polypeptide 1-like)	Z47087
RNA Polymerase III subunits	
RNA polymerase III subunit (RPC39)	U93869
RNA polymerase III subunit (RPC62)	U93867
RNA Elongation	
Elongation Factor I-Beta	X60489
Elongation Factor S-II	M81601
Elongation	
TCEA (110kD)	OMIM 601425
TCEB1	L34587
TCEB (18kD)	
TCEB1L	
TCEB3	L47345
TCEC (15kDa)	
TFIIS (Transcription Elongation Factor IIS)	601425
E2F1 (E2F Transcription Factor)	M96577
TFAP2A (Transcription Factor A2 Alpha)	X95694
TFCP2 (Transcription Factor CP2)	U01965
TFC12 (Transcription Factor I2)	M65209
PRKDC (Protein Kinase, DNA activated catalytic subunit)	U47077
Termination of RNA polymerization	
Factors that regulate RNA polymerization	
General factors	
SUPT6H	U46691
TFIIA gamma subunit	U14193
TFIIA delta	
TFIIB related factor hBRF (HBRF)	U75276
TFIIE Alpha Subunit	X63468
TFIIE Beta Subunit	X63469
TFIIF, Beta Subunit	X16901
GTF2F1 (TFIIF)	X64037
GTF2F2 (TFIIF)	X16901
General Transcription Factor IIIA	U20272
TFIIH(52 kD subunit of transcription factor)	Y07595

89

232/116

TFIIH(p89)	
TFIIH(p80)	
TFIIH(p62)	U07595
TFIIH(p44)	OMIM 601748
TFIIH(p34)	OMIM 601750
Transcription Factor IIf(General transcription factor IIF, polypeptide 1 (74kD subunit))	X64037
Specific factors required for polymerization of essential genes	
BTF 62 kDSubunit (Basic transcription factor 62 kD subunit)	M95809
CAMP-dependent transcription factor ATF-4	M86842
CCAAT box-binding transcription factor 1	X92857
CRM1(Negative regulator CRM1)	Y08614
Cyclic-AMP-dependent transcription factor ATF-1	X55544
GABPA(GA-binding protein transcription factor, alpha subunit (60kD))	U13044
ISGF-3(Signal transducer and activator of transcription 1- alpha/beta)	M97935
NFIX(Nuclear factor I/X (CCAAT-binding transcription factor))	L31881
NFYA(Nuclear transcription factor Y, alpha)	M59079
NTF97(Nuclear factor p97)	L38951
Nuclear factor I-B2 (NFIB2)	U85193
Nuclear factor NF45	U10323
Nuclear factor NF90	U10324
POU2F1(POU domain, class 2, transcription factor 1)	X13403
Sp2 transcription factor	M97190
TCF12(Transcription factor 12 (HTF4, helix-loop-helix transcription factors 4))	M83233
TCF3(Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47))	M31523
TCF6L1(Transcription factor 6-like 1)	M62810
TF P65(Transcription factor p65)	L19067
TFCOUP2(Transcription factor COUP 2 (a.k.a. ARP1))	X91504
Transcription factor IL-4 Stat	U16031
Transcription Factor S-II (Transcription factor S-II-related protein)	D50495
Transcription factor Stat5b	U48730
Transcription Factor	L06633
Transcription factor (CBFB)	L20298
RNA Processing Factors	
RNA splicing and other processing factors	
9G8 Splicing Factor (Pre-mRNA Splicing factor SRP20)	L22253
CC1.3(Splicing factor (CC1.3))	L10910
HnRNP F protein	L28010
HNRPA2B1(Heterogeneous nuclear ribonucleoproteins A2/B1)	M29065
HNRPG(Heterogeneous nuclear ribonucleoprotein G)	Z23064

90

232/116

HNRPK(Heterogeneous nuclear ribonucleoprotein K)	S74678
Pre-mRNA splicing factor helicase	D50487
Pre-mRNA splicing factor SF2, P33 subunit	M69040
Pre-mRNA splicing factor SRP20	L10838
Pre-mRNA splicing factor SRP75	L14076
PRP4(Serine/threonine-protein kinase PRP4)	U48736
PTB-Associated Splicing Factor	X16850
Ribonucleoprotein A'	X06347
Ribonucleoprotein A1	X13482
Ribonucleoprotein C1/C2	M15841
RNP Protein, L (Heterogeneous nuclear ribonucleoprotein L)	X16135
RNP-Specific C(U1 small nuclear ribonucleoprotein C )	X12517
SAP 145(Spliceosome associated protein )	U41371
SAP 61(Splicesomal protein)	U08815
SC35(Splicing factor)	L37368
SF3a120	X85237
SFRS2(Splicing factor, arginine/serine-rich 2)	M90104
SFRS5(Splicing factor, arginine/serine-rich 5)	AF020307
SFRS7(Splicing factor, arginine/serine-rich 7)	L41887
Small nuclear ribonucleoprotein SM D1	J03798
SnRNP core protein Sm D2	U15008
SnRNP core protein Sm D3	U15009
SNRP70(U1 snRNP 70K protein)	M22636
SNRPB(Small nuclear ribonucleoprotein polypeptides B and B1)	J04564
SNRPE(Small nuclear ribonucleoprotein polypeptide E)	M37716
SNRPN(Small nuclear ribonucleoprotein polypeptide N)	U41303
Splicing factor SF3a120	X85238
Splicing factor U2AF 35 kD subunit	M96982
Splicing factor U2AF 65 kD subunit	X64044
SRP30C(Pre-mRNA splicing factor SF2, p33 subunit)	U30825
SRP55-2(Pre-mRNA splicing factor SRP75)	U30828
Transcription factor BTEB	D31716
Transcription initiation factor TFIID 250 kD subunit	D90359
RNA polyadenylation and cleavage	
Cleavage and polyadenylation specificity factor	U37012
Cleavage stimulation factor, 3' pre-RNA, subunit 1, 50kD	L02547
Cleavage stimulation factor, 3' pre-RNA, subunit 3, 77kD	U15782
HNRNP Methyltransferase	D66904
PABPL1(Poly(A)-binding protein-like 1)	Y00345
Pap mRNA(Poly(A) Polymerase)	X76770
RNA unwinding	
RNA Helicase	

91

232/116

GU Protein (ATP-Dependent RNA helicase dead)	U41387
KIAA0224 Gene(Putative ATP-dependent RNA helicase)	D86977
RNA Helicase A	L13848
RNA Helicase P110	U50553
Ste13(Nuclear RNA Helicase)	U90426

RNA Degradation

RNA modification

RNA Transport

#### 6) Genes Required to Maintain Integrity and Function of Cellular and Subcellular Structures

##### 6.1 Genes Required to Move Proteins, Small Particles, and Other Ligands Across Membranes to Maintain their Concentration at Levels Compatible with Cell Growth or Survival

Genes required to form coated pits and vesicles

Clathrins

AP47(Clathrin Coat Assembly AP47 )	D38293
AP50(Clathrin Coat Assembly Protein AP50)	U36188
Cell Surface Protein (Clathrin Heavy Polypeptide-Like Protein )	X83545
Cltb(Clathrin Light Chain B)	M20470
Cltc (Clathrin Heavy Chain)	U41763

##### 6.2 Genes Required to Transmit Signals within Cells at Levels Compatible with Cell Growth or Survival

Genes required to transmit signals from membranes

Adenylate Cyclase

Adenylate Cyclase	D63481
Adenylate Cyclase, II	X74210
Adenylate Cyclase,IV	D25538

Genes required to transmit signals within cellular compartments

##### 6.3 Genes Required to Maintain Cellular Energy Stores at Levels Compatible with Cell Growth or Survival

Genes required to Produce ATP from catabolism of sugar

Genes required for glycolysis (anaerobic and aerobic)

Genes required for oxidative phosphorylation

Complex I

MTND1 (Subunit ND1)	OMIM 51600
MTND2 (Subunit ND2)	OMIM 51601
MTND3 (Subunit ND3)	OMIM 51602
MTND4 (Subunit ND4)	OMIM 51603
MTND4L (Subunit ND4L)	OMIM 51604
MTND5 (Subunit ND5)	OMIM 51605
MTND6 (Subunit ND6)	OMIM 51606

Complex II

Complex III

Cytochrome b subunit

Complex IV

92

232/116

CO1 (Cytochrome c Oxidase Subunit 1)	OMIM 516030
CO2 (Cytochrome c Oxidase Subunit 2)	AF035429
CO3 (Cytochrome c Oxidase Subunit 3)	
Complex V	
ATP Synthase Subunit ATPase 6	OMIM 516060
6.4 Genes Required to Transport or Dock Vesicles, Polypeptides or Other Solutes Moving Between Cellular Compartments at Rates and Levels Compatible with Cell Growth or Survival	
Transport to, from or within the cytoplasm	
Kinesins	
Kinesin Heavy Chain	X65873
Kinesin Light Chain	L04733
Syntaxin	
Syntaxin 1a	L37792
Syntaxin 1b	U07158
Syntaxin 3	U32315
Syntaxin 5a	U26648
Syntaxin 7	U77942
Transport to, from or within the endoplasmic reticulum	
CANX (Calnexin)	M94859
ER Lumen Protein 1	M88458
ER Lumen Protein 2	X55885
Ribophorin I	Y00281
Ribophorin II	Y00282
Signal recognition particle receptor	X06272
SRP Protein	U20998
TIM17 preprotein translocase	X97544
Transport to, from or within the Golgi apparatus	
Golgin-245	U31906
TGN46 (Trans-Golgi Network Integral Membrane Protein TGN38 Precursor )	X94333
Transport to, from or within the other membrane bound compartments	
Beta-Cop	X82103
Coatomer Beta' Subunit	X70476
Coatomer Delta Subunit	X81198
Gp36b Glycoprotein (Vesicular integral-membrane protein VIP36 precursor)	U10362
Homologue of yeast sec7	M85169
Protein transport protein SEC13 (Chromosome 3p25)	L09260
SEC14 (S. Cerevisiae)	D67029
Synaptic vesicle membrane protein VAT-1	U18009
Synaptobrevin-3	U64520
Synaptotagmin I	M55047
Transmembrane(COP-coated vesicle membrane protein p24 precursor)	X92098

Vacuolar-Type (Clathrin-coated vesicle/synaptic vesicle proton pump 116 kd subunit )	Z71460
Transport to, from or within the nucleus	
Nuclear membrane constituents	
140 kD Nucleolar phosphoprotein	D21262
Autoantigen p542	L38696
Export protein Rae1 (RAE1)	U84720
Heterogeneous nuclear ribonucleoprotein A1	X79536
Nuclear pore complex protein hnup153	Z25535
Nuclear pore complex protein NUP214	D14689
Nuclear pore glycoprotein p62	X58521
Nuclear Transport Factor 2	X07315
Nucleoporin 98 (NUP98)	U41815
NUP88	Y08612
Ribonucleoprotein A	M29063
Ribonucleoprotein B"	U23803
Nuclear envelope & pore constituents	
Karyopherin	
Importin Alpha Subunit	D89618
TRN (Transportin)	U70322
6.5 Genes Required to Maintain Cell Shape and Motility at Levels Compatible with Cell Growth or Survival	
Cell structure genes (Cytoskeleton)	
Actin	X04098
Beta-Contractin	X82207
Capping Protein Alpha	U03851
CFL1 (Cofilin, Non-Muscle Isoform)	X95404
Desmin	J03191
Dystrophin	U26743
Gelsolin	X04412
hOGG1(Myosin Light Chain Kinase)	AB000410
IC Heavy Chain	U31089
Itga2 (Integrin, Alpha 2 (CD49B, alpha 2 Subunit of VLA-2 receptor))	X17033
Itga3 (Integrin Alpha-3 Precursor)	M59911
Keratin 19	Y00503
Keratin, Type II	J00269
Lamin A	M13451
LBR(Lamin B Receptor)	L25931
Light Chain Alkali	M22920
MacMarcks mRNA	X70326
MAP1a (Microtubule-Associated Protein 1A)	U14577
MAP2(Microtubule-Associated Protein 2)	U01828
MEG1(Protein-Tyrosine Phosphatase MEG1)	X79510



94

232/116

Microtubule-Associated Protein TAU	J03778
Suppressor Of Tubulin STU2	X92474
TUBg (Tubulin Gamma Chain)	M61764
Tubulin Alpha-4 Chain	X06956
USH1b (Myosin II Heavy Chain)	U39226
Villin	X12901
Villin 2 (Ezrin)	J05021
Genes required for cell motility	
Actin genes	
Actin Depolymerizing	S65738
Capping (Actin Filament)	M94345
Myosin genes	
MYH9(Myosin, Heavy Polypeptide 9, Non-Muscle)	M31013
MYL5(Myosin Regulatory Light Chain 2)	L03785
Myosin Heavy Chain 95F	U90236
Myosin Heavy Chain IB	D63476
Myosin IB	U14391
Sh3p17(Myosin IC Heavy Chain)	U61166
Sh3p18(Myosin IC Heavy Chain)	U61167
K1AA0059(Dematrin:Actin-Bundling Protein)	D31883
TTN (Titin:Myosin Light Chain Kinase)	X69490
6.6 Genes Required to Eliminate, Transform, Sequester or Otherwise Regulate Levels of Endogenous Cellular Toxins or Waste Substances at Levels Compatible with Cell Growth or Survival	
Organelles that transform or sequester toxic or waste substances	
Vacuoles	
ATP6c(Vacuolar H <sup>+</sup> ATPase proton channel subunit)	M62762
Lysosomes	
ATP6a1 (ATPase, H <sup>+</sup> Transporting, Lysosomal (Vacuolar Proton Pump), Alpha Polypeptide, 70kD)	L09235
ATP6b1(ATPase, H <sup>+</sup> transporting, lysosomal (vacuolar proton pump), beta polypeptide, 56/58kD)	M25809
ATP6d(ATPase, H <sup>+</sup> transporting, lysosomal (vacuolar proton pump) 42kD)	X69151
ATP6e(ATPase, H <sup>+</sup> transporting, lysosomal (vacuolar proton pump) 31kD)	X71490
ATPase, H <sup>+</sup> transporting, lysosomal (vacuolar proton pump) 31kD	X76228
Free radical inactivation	
Superoxide Dismutase	X02317
Maintenance of cellular redox potential at levels compatible with cell survival	

**Conditionally essential genes**

As indicated in the Summary, some genes are conditionally essential, meaning that they are essential for cell survival or proliferation only in certain circumstances.

Most commonly such circumstances are related to changes in the environment,

5 such as changes in the concentration of specific constituents such as nutrients, administration of pharmaceuticals (drugs), or physical elements affecting the cell.

In many cases the changes in the environment may be induced as part of a treatment regiment for cancer such as the administration of drugs or ionizing radiation. In the presence of such specific environmental changes or therapies,

10 genes with are not normally essential for cell survival or proliferation become essential and, consequently, targets for therapy under the present invention.

Therapy with inhibitors of conditionally essential genes involves administration of the inhibitor together with a chemical or physical elements that causes the target gene to be essential for cell survival or proliferation. The use of allele specific

15 inhibitors in the current invention allows specific killing of cancer cells with such chemical or physical agent since the gene function that is essential for the survival of cells (in the presence of the chemical or physical agent) is inhibited in the cancer cell but not in the normal cell.

This strategy begins with the identification of heterozygous alleles of genes coding

20 for proteins that are conditionally essential for cell viability or growth due to change in the chemical or physical environment. In one aspect of this invention, the gene targets of this application are responsible for mediating cell response to changes in the environment. Such environmental alterations include, for example, changes in the concentration of naturally occurring constituents such as amino

25 acids, sugars, lipids and inorganic and organic ions, as well as larger molecules such as hormones or antibodies, or changes in the partial pressure of oxygen or other gasses. The absence of a specific constituent in the environment makes the genes that are involved in synthesizing that nutrient within the cell essential,

whereas if the constituent were present in the environment in sufficient quantities, such genes would not be essential. Alternatively, high concentrations of a specific constituent in the environment may make genes that are responsible for eliminating or detoxifying that constituent within the cell essential, whereas, if the constituent  
5 were absent or present in normal concentrations, such genes would not be essential. Changes thus may involve either an increase or a decrease in specific constituents of the environments including nutrients, inorganic, or organic materials.

In another aspect of this invention, the gene targets of this application are  
10 responsible for maintaining cell survival or proliferation in the presence of a drug or biological material. For example, a drug that inhibits one pathway for maintaining the level of a cellular constituent within levels required for cell survival or proliferation may make alternative pathways essential. In a specific embodiment, the inhibition of a synthetic pathway for a cellular constituent may make alternative  
15 synthetic pathways essential for cell survival or proliferation. Alternatively, a drug that is toxic to the cell will make genes that are involved in the elimination, degradation, or excretion of the drug from the cell essential for continued survival or proliferation. It will be evident to those skilled in the art that anything which inhibits the ability of a cell to survive in the presence of a specific drug that is  
20 designed to be cytostatic or cytotoxic, will sensitize that cell to the effects of the drug. A "chemosensitizing" agent is one that inhibits a function in the cell that is conditionally essential due to the administration of a chemotherapeutic drug.

In another aspect of this invention, the gene targets of this application are responsible for maintaining cell survival or proliferation in response to external  
25 physical forces including, but not limited to, electromagnetic radiation of various amplitudes and wavelengths, including ionizing and nonionizing radiation and heating or cooling. In the presence of ionizing radiation, for example, genes that are

involved in DNA repair may be essential that are not essential in the absence of the external physical force. An agent that inhibits functions in the cell that are essential due to the administration of ionizing radiation would be termed a "radiosensitizing" agent.

- 5 In each instance, treatment of cancer or noncancer proliferative diseases may be achieved by identifying genes that are conditionally essential in the presence of specific environmental, pharmacological, or physical factors, determining whether such genes are subject to loss of heterozygosity, identifying alternative alleles in these genes and developing allele specific inhibitors of alternative forms of the gene.
- 10 The administration of such an inhibitor to a patient who has two alternative forms of the gene in normal cells but only one in the cancer cell due to LOH, together with the environmental, pharmacological or physical factors will result in an antiproliferative effect or killing of the cancer cell.

- Different environmental, pharmacological, and physical changes in the environment
- 15 that result in homeostatic or compensatory responses in which genes that are not normally essential for cell survival or proliferation become essential are known in the art. These are described in the following Table 2.

**Table 2**

- |    |  |
|----|--|
| 1  | Changes in the concentration of constituent in the environment   |
| 20 | <input type="checkbox"/> Change in nutritional environment<br><input type="checkbox"/> Change in hormonal environment<br><input type="checkbox"/> Change in the immunological environment<br><input type="checkbox"/> Presence or accumulation of toxic materials<br><input type="checkbox"/> Change in partial pressure of oxygen |
| 25 | <input type="checkbox"/> Change in partial pressure of carbon dioxide.<br><input type="checkbox"/> Change in partial pressure of other gasses including nitrous oxide  |
| 2. | Administration of pharmaceuticals including small molecules, biologicals, nucleic acids, or antibodies.  |

3. Physical changes
  - ☐ Electromagnetic radiation
  - ☐ Ionizing radiation including Alpha particles, Beta particles, Gamma radiation
  - 5 ☐ Non-ionizing radiation including infrared radiation, microwave radiation, other wavelengths
  - ☐ Temperature

When LOH results in a difference in normal cell genotype vs. cancer cell genotype that affects a locus encoding a product affecting the cells' ability to survive in the

10 presence of an environmental change, a pharmaceutical or biological agent, or a physical factor, there is an opportunity to exploit a therapeutic window between cancer cells and normal cells. Below we describe specific examples of genes that (1) affect cell responses to altered environments, (2) are located on chromosomes that undergo LOH in cancer and (3) exist in two or more variant forms. These examples

15 have been selected to illustrate how the therapeutic strategy described in this application would work with a variety of different alterations in chemical or physical environment. Example 43 describes a gene (Dihydropyrimidine Dehydrogenase) that mediates response to an altered chemical environment (presence of the toxic chemical 5-floxuridine) by specifically transforming the chemical to an inactive

20 metabolite. Example 39 describes a gene (Methylguanine methyltransferase) that mediates response to an altered chemical environment (presence of toxic chemicals such as nitrosourea or other alkylating agents) by removing methyl or alkyl adducts to DNA, the principal toxic lesion of these agents. Example 44 describes a set of genes (Fanconi Anemia genes A,B,C,D,E,F,G and H) which mediate response to an

25 altered chemical environment (presence of chemicals which cause DNA crosslinking, such as diepoxybutane, mitomycin C and cisplatin) by repairing the crosslinks. Example 48 describes a set of genes (the DNA Dependent Protein Kinase Complex, including the DNA Dependent Protein Kinase catalytic subunit (DNA-PKcs), the DNA binding component (called Ku), made up of Ku-70 and

30 Ku-86 kDa subunits, and the Ku-86 related protein Karp-1) that mediates repair of

double stranded DNA breaks, such as occurs after x-irradiation. Example 45 describes a gene (asparagine synthase) that mediates response to an altered nutritional environment (absence of extracellular asparagine) which can be produced by an enzyme such as asparaginase, which hydrolyzes serum asparagine. Example 5 49 describes the Ataxia Telangiectasia gene, which is involved in response to ionizing radiation and radiomimetic chemicals. Other detailed examples include methionine synthase (Ex. 46) and methylthioadenosine phosphorylase (Ex. 47). Other examples include Poly (ADP) Ribose Polymerase (PARP), Glutathione-S-Transferase pi (GST-pi), NF-kappa B, Abl Kinase, 3-alkylguanine alkyltransferase, 10 N-methylpurine DNA glycosylase (hydrolyzes the deoxyribose N-glycosidic bond to excise 3-methyladenine and 7-methylguanine from alkylating agent-damaged DNA polymers), OGG-1, MDR-1.

The table below presents exemplary categories and exemplary specific genes along with the type of conditions which render the gene essential.

### 15 **Table 3: Categories of Conditionally Essential Genes**

#### **Genes and proteins vital for cell survival or proliferation in the presence of an altered chemical or physical environment**

##### **I. Genes required for adaptation to changes in the chemical environment**

20

##### **1. Adaptation to altered concentration of a naturally occurring small molecule**

##### **A. Increased concentration of a naturally occurring small molecule**

25

##### **i. Increased levels of amino acids**

##### **1. Targets: amino acid degradation pathways**

100

232/116

Increased intracellular levels of amino acids can damage cells. One cause of such increased levels is failure to properly degrade amino acids into simpler compounds. Therefore an amino acid catabolizing enzyme can be a conditionally essential gene, particularly in the presence of elevated levels of the twenty amino acids commonly used in protein synthesis. Amino acid catabolic pathways are well described in textbooks and in the scientific literature.

ii. Increased levels of sugars or starches

2. Targets: mono, di and polysaccharide metabolic pathways

**Galactose-1-phosphate uridylyltransferase**

**Galactose kinase**

**UDPgaltactose-4-epimerase**

Increased intracellular levels of sugars or starches can damage cells. One cause of increased levels is failure to properly degrade starches into simple compounds, as exemplified by diseases of impaired polysaccharide metabolism. Therefore a polysaccharide catabolizing enzyme can be a conditionally essential gene, specifically in the presence of elevated levels of particular polysaccharides. A second mechanism of damage arises in the context of impaired sugar metabolism. Thus enzymes that degrade sugars or starches to simpler compounds may be conditionally essential for cell health and consequently cell proliferation. An example is the enzymes of the Leloir pathway of galactose metabolism. Mutant copies of these proteins make cells conditionally sensitive to elevated concentrations of galactose. Thus enzymes that degrade sugars or starches to simpler compounds may be conditionally essential for cell proliferation.

iii. Increased levels of vitamins

B. Decreased concentration of a naturally occurring small molecule

i. Decreased levels of amino acids

1. Targets: amino acid transporters

Decreased intracellular levels of amino acids can impair protein synthesis and thereby slow or arrest cell division. One cause of such decreased levels is impairment of cellular uptake of amino acids, particularly amino acids that the cell is not actively synthesizing, whether essential (e.g. methionine) or nonessential (e.g. asparagine; see examples). Cells have a variety of mechanisms for amino acid uptake, including membrane anchored transporters. In the presence of decreased extracellular levels of amino acids the protein and other constituents of these transporters become conditionally more essential.

2. Targets: amino acid biosynthetic machinery

a. Essential amino acids

**Methionine Synthase**, essential for responding to decreased extracellular methionine. (GenBank U73338)

b. Non-essential amino acid biosynthesis

**Asparagine Synthase**, essential for responding to decreased extracellular asparagine. (GenBank M27396)

**Glutamine Synthetase**, essential for responding to decreased extracellular glutamine. (GenBank Y00387)



102

232/116

Decreased intracellular levels of amino acids can impair protein synthesis and thereby slow or arrest cell division. One cause of such decreased levels is impairment of amino acid biosynthesis, particularly amino acids that the cell is not actively synthesizing, whether essential (e.g. methionine) or nonessential (e.g. asparagine; see examples). Cells have a variety of well described biochemical pathways for biosynthesis of the 20 amino acids commonly used in proteins. These biosynthetic enzymes can be conditionally essential in the absence of adequate intracellular levels of amino acids. Specific examples of such conditionally essential genes are described in the Examples. However, other enzymes which catalyze reactions important for maintaining levels of amino acids adequate for protein synthesis in the presence of decreased extracellular concentrations are also useful.

### 3. Targets: transaminases

In the presence of decreased extracellular levels of amino acids cells must increase intracellular mechanisms for amino acid biosynthesis. One such mechanism is transfer of amino groups from nonessential to essential amino acids to compensate for insufficient quantities of essential amino acids. These reactions are catalyzed by transaminases, which therefore can become conditionally essential in environments characterized by decreased levels of extracellular amino acids.

5

### ii. Decreased levels of sugars

#### 1. Targets: sugar transporters

#### 2. Targets: sugar metabolism machinery

10

103

232/116

Increased intracellular levels of sugars or starches can damage cells. One cause of such increased levels is failure to properly degrade starches into simple compounds, as exemplified by diseases of impaired polysaccharide metabolism. Therefore a sugar or poly-saccharide catabolizing enzyme can be a conditionally essential gene in the presence of elevated levels of particular sugars or polysaccharides.

2. Adaptation to presence of non-naturally occurring molecules

5 A. Elimination of non-naturally occurring molecules

i. Elimination by export

10

**Multidrug resistance gene/P glycoprotein (MDR1)**

(GenBank AF016535)

**Multidrug resistance associated proteins 1-5 (MRPs)**

(GenBank L05628)

Cells have evolved specific mechanisms to export a variety of chemicals, including nonnatural chemicals such as cytotoxic drugs. MDR1 and MRP are exemplary ATP-dependent transmembrane drug-exporting pumps. Deficiency of these pumps is associated with increased sensitivity to a variety of cytotoxic drugs in vitro and in vivo. For example, mice lacking functional MRP are hypersensitive to the drug etoposide. Thus these pumps are important for cell survival in the presence of a variety of toxic drugs. Polymorphisms have been reported in MDR1 at amino acids 893 and 999. MDR also maps to a region of chromosome 7 which is frequently affected by LOH in prostate, ovarian breast and other cancers.

15

**Multispecific organic anion transporters (MOATs)**

**Other drug export proteins**

ii. Elimination by metabolic transformation

20

1. Specific metabolic transformation of drugs

a. Inactivation of bleomycin

5

**Bleomycin hydrolase (GenBank U14426)**

Bleomycin hydrolase was discovered through its ability to detoxify the anticancer glycopeptide bleomycin. Cells lacking bleomycin hydrolase are highly susceptible to bleomycin toxicity (for example pulmonary fibrosis) thus the gene is conditionally essential for cell growth and survival in the presence of bleomycin. Bleomycin hydrolase is a member of the cysteine protease papain superfamily. The protein is expressed in all tissues surveyed. The crystal structure of the closely related yeast bleomycin hydrolase has been determined. A common A/G polymorphism has been described at nucleotide 1450 of the bleomycin hydrolase gene. It results in an isoleucine-valine variance at amino acid 443, part of the oligomerization domain of the homotetrameric enzyme. The Bleomycin hydrolase gene has been mapped to the proximal long arm of chromosome 17 (17q11.2), a site of frequent LOH in commonly occurring epithelial cancers such as breast and ovarian cancer.

b. Inactivation of pyrimidine analogs including 5-fluorouracil (5-FU) and 5-fluorouridine.

10

**Dihydropyrimidine Dehydrogenase (DPD)**

$\beta$ -ureidopropionase

$\beta$ -alanine synthetase

DPD is described in the examples. The other two enzymes are responsible for the further metabolism of dihydro-5-fluorouracil, the metabolic product of DPD. In the absence of these enzymes toxic metabolites of 5-FU accumulate in cells.

15

c. Inactivation of pyrimidine analogs including cytosine arabinoside and 5-azacytidine.

105

232/116

**Cytidine deaminase**

Cytidine deaminase (CDA) catalyzes hydrolytic deamination of cytidine or deoxycytidine. It can also deaminate cytotoxic cytosine nucleotide analogs such as cytosine arabinoside, rendering them nontoxic. Resistance to the cytotoxic effects of these drugs has been reported associated with increased expression of the CDA gene. Thus CDA is a conditionally essential gene in the presence of cytotoxic cytosine nucleotide analogs.

- 5 d. Inactivation of thiopurine drugs, including 6-mercaptopurine, 6-thioguanine and azathioprine.

**Thiopurine methyltransferase (GenBank U12387)**

- 10 e. Inactivation or transformation of other drugs including, but not limited to, purine analogs, folate analogs, topoisomerase inhibitors and tubulin acting drugs via specific enzymatic modification.

15 2. General metabolic transformation of drugs

a. Cytochrome P450 system.

**CYP1**

CYP1A1 (GenBank K03191)

CYP1A2 (GenBank M55053)

20 **CYP2**

CYP2A6 (GenBank U33317)

CYP2A7

CYP2B6

CYP2B7

25 CYP2C8

CYP2C9 (OMIM 601130)

CYP2C17

CYP2C18

CYP2C19 (OMIM 124020)

30 CYP2D6 (OMIM 124030)

CYP2E1 (OMIM 124040)

CYP2F1

**CYP3**

CYP3A3

35 CYP3A4 (GenBank D00003)

CYP3A5

CYP3A7

106

232/116

5

CYP4  
CYP4B1  
CYP7  
CYP11  
CYP17  
CYP19  
CYP21  
CYP27

10

The cytochrome P450s are a large gene family whose members metabolically transform and inactivate a wide variety of drugs, including cytotoxic drugs. Wide variation in P450 protein expression has been described, including null alleles. For example cytochrome P450 2D6 may be involved in the metabolism of ~25% of all drugs. Between 5 and 10% of all caucasians are homozygous for completely inactive alleles of P450 2D6. In the presence of a toxic drug the P450 enzyme responsible for metabolizing the drug may be conditionally essential. For example, acute liver failure has been reported in a patient treated with cyclophosphamide who was homozygous for the deficient CYP 2D6B allele. Liver failure was due to accumulations of a hepatotoxic 4-hydroxylated cyclophosphamide metabolite.

b. N-acetyltransferases

15

c. Glucuronyltransferases

d. Glutathione transferases

20

**Glutathione transferase alpha** (GenBank AF020919)  
**Glutathione transferase theta** (OMIM 600436 & 600437)  
**Glutathione transferase mu** (OMIM 138350, 138380,  
138380, 138333 & 138385)  
**Glutathione transferase pi** (GenBank X65032)

A large number of drugs are are biotransformed into electrophilic intermediary compounds which are potentially harmful to cell constituents unless rendered harmless by conjugation with glutathione. Thus proteins of the GST system are conditionally essential for cell survival.

B. Repair or prevention of damage by non-naturally occurring molecules

5 i. Repair or prevention of damage by molecules that react with nucleic acids

1. Molecules that add alkyl or other groups to DNA

10 a. Targets: genes & gene products involved in repair of alkylating agent damage

**Methylguanine Methyltransferase (MGMT) (GenBank M29971)**

15 **3-alkylguanine alkyltransferase**

**3-methyladenine DNA glycosylase (GenBank M74905)**

MGMT is described in the examples. hOGG1 is a DNA glycosylase with associated lyase activity that excises this adduct and introduces a strand break. Cells lacking this protein are deficient in repair of oxidative damage and have high mutation rates. In conditions of high oxidative damage, including cellular aerobic metabolism, ionizing radiation and some chemotherapy drugs the hOGG1 gene would be conditionally essential for DNA repair. The human OGG1 gene maps to chromosome 3p25, a region of high frequency LOH in lung, kidney, head and neck and other cancers. Homozygous mutant mouse cells lacking 3-methyladenine DNA glycosylase have increased sensitivity to alkylation induced chromosome damage and cell killing.

20 2. Molecules that induce single or double stranded DNA breaks (also relevant to survival in the presence of ionizing radiation; see below)

a. Targets: genes & gene products involved in repair of double stranded DNA breaks

25

**DNA Dependent Protein Kinase (DNA-PK) and subunits**  
**Catalytic subunit of DNA-PK (GenBank U47077)**  
**DNA binding subunit of DNA-PK (Ku subunit)**  
**Ku-70 subunit (GenBank J04611)**

**Ku-86 subunit (OMIM 194364/GenBank AF039597)**

**KARP-1**

**Poly (ADP-ribose) polymerase (PARP) (GenBank M32721)**

- 5                   b. Targets: genes & gene products that repair DNA cross-links induced by molecules such as Mitomycin C or diepoxybutane

- 10                   **Fanconi Anemia genes**  
                       **Fanconi Anemia A gene** (GenBank X99226)  
                       **Fanconi Anemia B gene**  
                       **Fanconi Anemia C gene** (GenBank X66894)  
                       **Fanconi Anemia D gene**  
                       **Fanconi Anemia E gene**  
 15                   **Fanconi Anemia F gene**  
                       **Fanconi Anemia G gene**  
                       **Fanconi Anemia H gene**

- 20                   4. Targets: genes & gene products required for repair of DNA damage caused by drugs such as, for example, 4-nitroquinoline -1-oxide, bromobenz(a)anthracene, benz(a)anthracene epoxide, 1-nitorpyridine-1-oxide, acetylaminofluorine and aromatic amides, benz(a)pyrene.

- 25                   a. Nucleotide excision repair system  
                       **ERCC-1** (GenBank M13194)  
                       **ERCC2/XPD** (GenBank X52222)  
                       **ERCC3/XPB** (GenBank M31899)  
                       **ERCC4** (OMIM 133520)  
 30                   **ERCC5** (GenBank L20046)  
                       **ERCC6** (GenBank L04791)

- b. Other DNA repair genes  
                       **XPA** (GenBank D14533)  
 35                   **XPC** (GenBank D21090)  
                       **XPE** (GenBank U18300)  
                       **HHR23A** (GenBank U21235)  
                       **HHR23B** (GenBank D21090)  
                       **Uracil glycosylase** (GenBank X52486)  
 40                   **3-methyladenine DNA glycosylase** (GenBank M74905)

ii. Repair of damage by chemicals that interact with proteins

iii. Repair of damage by chemicals that interact with membranes

109

232/116

### 1. Free radical damage

5           iv. Adaptation to molecules that alter the cellular redox state (such as  
pyrrolidinedithiocarbamate)

### 3. Adaptation to change in nutritional environment

10           A. Decreased levels of nutrients.

B. Increased levels of nutrients.

### 4. Change in hormonal environment

15           A. Decreased levels of hormones.

B. Increased levels of hormones.

### 20           5. Change in the immunological environment

A. Introduction of new immune molecules (antibodies or antibody  
fragments)

25           B. Introduction of immune regulatory molecules

### **Fanconi anemia C** **NF-kappa B (GenBank M58603)**

Cells lacking the Fanconi anemia C gene have been shown hypersensitive to interferon gamma in vitro. Cells lacking the RelA/p65 subunit of NF kappa B are essential for preventing Tumor Necrosis Factor alpha induced cell death. Other Fanconi anemia genes or other proteins of the NF-Kappa B system and its regulators, for example I kappa B, may also mediate sensitivity to immune system molecules, for example interferons, interleukins or TNF.

30

## II. Changes in physical environment

### 1. Repair of damage caused by electromagnetic radiation



110

232/116

A. Repair of damage caused by ionizing radiation (Alpha particles, Beta particles, Gamma radiation)

i. DNA-PK constituents (see above)

ii. Other proteins that repair DNA damage created by DNA-PK

5 XRCC4 (GenBank U40622)

XRCC5/Ku80 (OMIM 194364)

XRCC6

XRCC7 (GenBank L27425)

iii. Other proteins that repair or protect from DNA damage

10 Glutathione-S-transferase (alpha, theta, mu and pi proteins)

Transfection of an exogenous Glutathione-S-transferase pi (GST-pi) gene is partially protective of cells treated with ionizing radiation. Thus GST activity is conditionally essential for cells exposed to ionizing radiation. Similarly, any protein that is essential for the repair of radiation induced damage or for protection of cells from radiation induced damage is a conditionally essential gene. GST activity can also affect radiation sensitivity in the presence of electron affinic drugs such as the nitroimidazoles.

15 I-kappa B alpha (GenBank M69043)

Increased expression of exogenous I kappa B-alpha, an inhibitor of NF-kappa B, increases cell sensitivity to ionizing radiation. Thus is conditionally essential for cells exposed to ionizing radiation. Other proteins of the NF kappa B pathway that affect radiosensitivity are likewise conditionally essential in the presence of ionizing radiation.

B. Non-ionizing radiation

20 i. infrared radiation

ii. ultra high frequency electromagnetic radiation (UHF)

**Glutathione S transferase system** (see genes listed above)

111

232/116

UHF electromagnetic radiation of 434 Mhz will change resonance of the glutathione cycle resulting in thiol depletion which increases radiosensitivity. UHF is therefore a radiosensitizing treatment, contingent on the status of the glutathione system.

### iii. Other wavelenths of electromagnetic radiation

#### 5 2. Temperature

##### A. Heating

##### 1. Heat shock proteins

**HSP70** (OMIM 138120)

10 **HSP27** (GenBank X54079)

##### B. Cooling

##### 2. Cold sensitive proteins

#### 15 3. Change in redox environment, including change in partial pressure of gasses

##### A. Change in partial pressure of oxygen

##### i. Repair of damage from reactive oxygen species

**8-oxoguanine DNA glycosylase (hOGG1)** (GenBank  
20 U96710)

The major mutagenic lesion caused by exposure to reactive oxygen species is 8-oxoguanine. hOGG1 is a DNA glycosylase with associated lyase activity that excises this adduct and introduces a strand break. Cells lacking this protein are deficient in repair of oxidative damage and have high mutation rates. In conditions of high oxidative damage, including cellular aerobic metabolism, ionizing radiation and some chemotherapy drugs the hOGG1 gene would be conditionally essential for DNA repair. The human OGG1 gene maps to chromosome 3p25, a region of high frequency LOH in lung, kidney, head and neck and other cancers.

25 Fanconi anemia genes (see above for list of 8 FA complementation groups; FA genes also mediate sensitivity to oxygen)

B. Change in partial pressure of carbon dioxide.

C. Change in partial pressure of other gases.

5

In addition to being hypersensitive to ionizing radiation Ataxia-Telangiectasia cells are hypersensitive to the nitric oxide donor S-nitrosoglutathione (GSNO), as are cells from some radiosensitive individuals without ataxia. GSNO induces dose-dependent DNA strand breakage; cell killing appears to be associated with formation of nitrite as the ultimate oxidation product of nitric oxide. Any protein important for response to damage induced by a dissolved gas is a conditionally essential gene in this category.

### III. Identification of variances and alternative alleles.

A target gene of this invention must occur as alternative alleles in the population;  
10 that is, the DNA sequence variance should either affect the gene sequence, RNA sequence, or protein sequence of the gene or its gene products, which would facilitate the design of inhibitors of the protein product, or be a base difference anywhere within the genomic DNA sequence, including the promoter or intron regions. Such DNA sequence variance can be exploited to design inhibitors of  
15 transcription or translation which distinguish between two allelic forms of the targeted gene. Sequence variants that do not alter protein sequence can be targeted, for example, with antisense oligonucleotides or ribozymes.

The most elementary genetic variant, which is common in mammalian genomes, is the single nucleotide substitution. It has been estimated that the comparison of  
20 haploid genomes will reveal this type of variant every 300 to 500 nucleotides (Cooper, *et al.*, *Human Genetics*, 69:201:205 (1985)).

Sequence variances are identified by testing DNA from multiple individuals from

- the population(s) to determine whether the DNA sequence for the target gene differs in different individuals. Many different methods for identifying gene sequence variances are known in the art, several of which are described in detail in the Examples noted below. These include, but are not limited to: (1) sequencing using
- 5 methods such as Sanger sequencing which is commonly performed using automated methods (Example 37); (2) Single Strand Conformation Polymorphism (Example 28); (3) DGGE (Example 36); (4) Computational methods (Example 30); (5) Chemical cleavage, (6) HPLC; (7) Enzymatic Mutation Detection, (Example 29); (8) Hybridization; (9) Hybridization arrays; and (10) Mass spectroscopy.
- 10 Often combinations of these methods are used. For example, methods such as SSCP, DGGE, or HPLC are useful in identifying whether amplified gene segments from two individuals are identical or contain a variance. These methods do not identify the location of the variant site within the linear sequence of the amplified gene segment, nor do these methods identify the specific nature of the variance,
- 15 namely the alternative bases within the variant site. Methods such as Enzymatic Mutation Detection determines where the variant site is located within the sequence, but not the specific variance. Methods such as mass spectroscopy identify the specific variance, but not its location within the segment. Methods such as sequencing, computational analysis, and hybridization arrays can determine the
- 20 location of the variance and specific sequence of the variance within the segment. In addition, methods such as SSCP, DGGE, EMD, and chemical cleavage are useful for determining alleles containing more than one variant site, if such sites occur within a single amplified gene segment. For the purpose of this invention, methods have been used to identify novel variant sites within genes that are essential for cell
- 25 survival or proliferation. With the above methods, the presence and type of variance are preferably confirmed, such as by sequencing PCR amplification products extending through the identified variance site.

#### IV. Loss of Hertozygosity

Essential genes which are located in chromosomal regions which frequently undergo LOH in a tumor or other disease or condition provide advantageous targets, as the LOH of the chromosomal region indicates that the particular gene will also  
5 undergo LOH at similar high frequency. Also, essential genes which undergo LOH at high frequencies in a particular tumor, or in a range of tumor types provide advantageous targets, as a large number of patients will be potentially treatable due to the LOH of a particular essential gene.

Cancer cells, or more broadly cells associated with certain other proliferative  
10 conditions, are generally genetically different from normal somatic cells as a result of partial or complete chromosome loss, called loss of heterozygosity (LOH), which occurs at the earliest stages of these disorders. In cancer, as a result of such early chromosome loss, all the tumor cells in an individual exhibit the same pattern of LOH since the cancer results from clonal expansion of the progenitor cell with  
15 LOH. Losses of genes in LOH range from less than 5% of a chromosome, to loss of a chromosome arm, to loss of an entire chromosome. Generally only one chromosome copy is lost, making cancer cells partially hemizygous - *i.e.*, they have only one allele of many genes. As a result of such allele loss, only the single remaining allele will be available to be expressed. Such loss of heterozygosity and  
20 other losses of genetic material in cancers is described in a variety of references, for example in Mitelman, F., Catalog of Chromosome Aberrations in Cancer, New York: Liss (1988); and Seizinger, *et al.*, "Report of the committee on chromosome and gene loss in neoplasia," *Cytogenet. Cell Genetics*, 58:1080-1096 (1991). A review of many published studies of LOH in cancer cells is provided in Lasko,  
25 Cavenee, and Nordenskjold, "Loss of Constitutional Heterozygosity in Human Cancer," *Ann. Rev. Genetics*, 25:281-314 (1991).

There is considered to be a causal relationship between LOH and the origin of

cancer or other proliferative disorders. Loss of heterozygosity commonly involves chromosomes and chromosome segment that contain at least one tumor suppressor gene in addition to many other genes that may not have any function associated with cancer but are coincidentally located in the same region of the chromosome, measured in physical distance or genetic distance, as the tumor suppressor gene. Tumor suppressor genes generally regulate cell proliferation or are involved in initiating programmed cell death when threshold level of damage occurs to the cell. The loss of tumor suppressor gene function is believed to confer a growth advantage to cells undergoing LOH, because it allows them to evade these negative growth regulatory events. It is the loss of tumor suppressor genes, and the proliferative advantage associated with loss of tumor suppressor functions, that drives allele loss or loss of heterozygosity. Loss of tumor suppressor gene function requires inactivation of both gene copies. Inactivation is usually due to the presence of mutations on one gene copy and partial or complete loss of the chromosome, or chromosome region, containing the other gene copy. (Lasko et al., 1991, *Annu. Rev. Genet.* 25:281-314)

Several tumor suppressor genes have been cloned. They include, for example, TP53 on chromosome arm 17p, BRCA1 on 17q, RB and BRCA2 on 13q, APC on 5q, DCC on 18q, VHL on 3p, and p16<sup>INK4</sup>/MTS1 on 9p. Many other, as yet uncloned, tumor suppressor genes are believed to exist based on LOH data; research groups are currently working to identify new tumor suppressor genes at more than a dozen genomic regions characterized by high LOH in cancer cells, including generating detailed LOH maps which provide LOH information useful for this invention due to the ability to identify essential genes which map to these regions of LOH. While there is an extensive literature considering tumor suppressor genes as potential targets for anti-cancer therapy, these genes are, in general, not candidates for antiproliferative therapy under the present invention because most tumor suppressor genes are not essential for cell proliferation or survival. To the contrary,

it is the loss of tumor suppressor genes that enables the abnormal proliferation and survival of cancer cells.

The pattern of LOH for a particular cancer or tumor or other proliferative disorder is not merely random. Often, there is a characteristic pattern for each major cancer  
5 type. Certain regions, including segments of chromosomes 3, 9, 11, 13, and 17, are frequently lost in most major cancer types. Other regions, such as on chromosomes 1, 3, 5, 6, 7, 8, 9, 11, 13, 16, 17, 18, and 22, exhibit high frequency LOH in selected cancers. It is believed that the characteristic LOH patterns of different cancers reflects the location(s) of tumor suppressor genes related to the development of the  
10 particular cancer or cancer type. Thus, essential genes located in regions which are characteristically associated with LOH for a particular cancer, or other tumor are particularly advantageous targets for inhibitors useful for treatment of that cancer or tumor because such genes will also characteristically undergo LOH at high frequency. The fact that certain cancers predictably undergo LOH in specific  
15 regions of the genome, and that LOH occurs before the clonal expansion of cancers in precancerous, abnormally proliferating tissue is potentially useful for preventing cancer with allele specific inhibitors of essential genes.

The treatment method described herein is applicable to proliferative disorders in which clonal proliferation occurs and in which the proliferating cells commonly  
20 undergo LOH. Another example of a disorder which has been characterized as a proliferative disorder is inflammatory pannus in arthritic joints. The demonstration of LOH associated with such a disorder will indicate that the allele specific treatment would be appropriate for the disorder. For the application of the general allele specific inhibition strategy to such conditions (*e.g.*, selection of target gene  
25 and variance, identification of inhibitors, selection of composition and administration method appropriate for the condition and the inhibitor), the cells associated with the condition correspond with the tumor, *e.g.*, cancer cells, for the

methods described in the Summary above.

LOH has been described for such polyclonal or oligoclonal disease conditions, in particular for atherosclerosis (arteriosclerosis), for example in Hatzistamou et al., 1996, *Biochem. Biophys. Res. Comm.* 225:186-190. Using a limited set of markers  
5 located on 18 chromosomal arms (one marker per arm), it was found that 23% of atherosclerotic plaques exhibited LOH for at least one marker. This does not necessarily represent the maximum fraction of plaques which could potentially be treated with allele specific inhibitors because the study did not attempt to determine the sites of maximum LOH on each arm. LOH which is partial arm LOH not  
10 affecting the particular marker for that arm was not detected. In general, fine scale LOH studies (using closely spaced markers) have revealed more sites of high frequency LOH than coarser scale studies.

The LOH for alleles of essential genes in cancers forms the basis for the anticancer therapeutic strategy described in Housman, *supra*. When one allele of the essential  
15 gene is lost from the patient's cancer cells, the retained allele can be targeted with an allele specific inhibitor. Such an inhibitor will kill, or reduce or prevent the growth of cancer cells by abolishing the function of an essential gene. Normal cells, which retain both uninhibited and inhibited alleles, will survive or grow due to the expression of the uninhibited allele. This is clearly indicated because tumor cells  
20 having only one allelic form (after LOH) thrive, thus, normal cells will also function normally with one of two allelic forms inhibited.

A large number of high frequency LOH regions are identified in Fig. 5. If not previously known, this correlation can be determined routinely for one or more tumor types by mapping of essential genes to chromosomal regions which have  
25 been identified as having high frequency LOH, or by identifying essential genes which map to locations near markers which have been identified as undergoing high



frequency LOH in a tumor. As previously described, the LOH of a marker near an essential gene, or the bracketing of an essential gene by two markers which undergo LOH, is strongly indicative that the essential gene also undergoes LOH at a similar frequency.

5 **TABLE 4**

**Loss of Heterozygosity in Human Solid Tumors By Chromosome Arm**

<u>Chromosome Region</u>	<u>Tumor Type</u>	<u>Chromosome Region</u>	<u>Tumor Type</u>
1p	Breast carcinoma	2	Uveal melanoma
10	Cutaneous melanoma (metastatic)		
	Medullary thyroid carcinoma: MEN2A		
15	Neuroblastoma		
	Pheochromocytoma: MEN2A sporadic		
1q	Breast carcinoma		
	Gastric adenocarcinoma		
3p	Breast carcinoma	4q	Hepatocellular carcinoma
20	Cervical carcinoma		
	Lung cancer: small carcinoma non-small cell carcinoma		
25	large cell carcinoma squamous cell carcinoma		
	adenocarcinoma		
30	Ovarian carcinoma		
	Renal cell carcinoma: familial sporadic		
	Testicular carcinoma		

119

232/116

5q	Colorectal carcinoma Hepatocellular carcinoma	6q	Ovarian carcinoma Primitive neuroectodermal tumor Renal cell carcinoma Testicular teratocarcinoma
9p 9q	Glioma Bladder carcinoma	10 10q	Glioblastoma multiforme Hepatocellular carcinoma Prostate cancer
5 10 15 20 25	11p Adrenal adenoma Adrenocortical carcinoma Bladder carcinoma Breast carcinoma Embryonal rhabdomyosarcoma Hepatoblastoma Hepatocellular carcinoma Lung cancer: squamous cell carcinoma large cell carcinoma adenocarcinoma Ovarian carcinoma Pancreatic cancer Parathyroid tumors Pheochromocytoma Skin cancer squamous cell carcinoma basal cell carcinoma Testicular cancer Wilms tumor 11q Insulinoma Parathyroid tumors	12q	Gastric adenocarcinoma

120

232/116

<p>5</p> <p>13q Adrenocortical adenoma Breast carcinoma Gastric carcinoma Hepatocellular carcinoma Lung cancer:     small cell carcinoma Neuroblastoma Osteosarcoma Retinoblastoma</p>	<p>14 Colorectal carcinoma 14q Neuroblastoma</p>
<p>10</p> <p>16 Breast carcinoma 16q Breast carcinoma Hepatocellular carcinoma Primitive neuroectodermal tumor 15 Prostate cancer</p>	<p>17p Adrenocortical adenoma Astrocytoma Bladder carcinoma Breast carcinoma Colorectal carcinoma Lung cancer:     small cell carcinoma     squamous cell carcinoma     adenocarcinoma Medulloblastoma Neurofibrosarcoma: NF1 Osteosarcoma Ovarian carcinoma Primitive neuroectodermal tumor Rhabdomyosarcoma 17q Breast carcinoma Neurofibroma: NF1</p>
<p>18 Renal cell carcinoma 18q Breast carcinoma Colorectal carcinoma</p>	<p>22q Acoustic neurinoma Colorectal carcinoma Ependymoma Meningioma Neurofibroma</p>

V. Use of variance-specific inhibitors of essential genes to treat non-malignant,  
20 proliferative conditions.

It was found in the present invention that noncancer proliferative disorders could also be targeted using such an allele specific strategy. Such conditions include, but are not limited to atherosclerotic plaques, abnormal tissue in arthritic joints, including pannus, benign tumors such as leiomyomas and meningiomas, and hyperplastic conditions such as benign prostatic hyperplasia. For most of these conditions there is evidence of a mono- or oligoclonal origin and evidence of LOH. Such evidence includes the following:

- A recent study (Hatzistamou, J., Kiaris, H., Ergazaki, M., et al. (1996) Loss of heterozygosity and microsatellite instability in human atherosclerotic plaques. *Biochemical and Biophysical Research Communications* 225: 186-190.) demonstrated that allele loss occurs in atheromatous plaques, which have long been viewed as benign neoplastic proliferations by some investigators (Benditt, E.P. and J.M. Benditt (1973) Evidence for a monoclonal origin of human atherosclerotic plaque. *Proc. Natl. Acad. Sci. U. S. A.* 70: 1753-7). Each atheromatous plaque constitutes a separate independently arising primary lesion. Consequently, allele loss in individual atherosclerotic plaques will differ, with, for example, allele A of a hypothetical essential gene lost in some plaques and allele A' in others. An inhibitor of allele A would be expected to kill (or arrest growth of) only about half of all the plaques with allele loss at the hypothetical locus - those plaques hemizygous for A. To kill the other half of the plaques with allele loss at the target locus would require an inhibitor of A'. Simultaneous use of inhibitors of A and A' would be highly toxic to diploid normal cells. However serial use of an inhibitor directed to allele A followed by an inhibitor directed to A' (perhaps repeating treatment for several cycles, or even indefinitely) would alternately abolish essential gene function in one half of all haploid plaque cells and then the other half, leading eventually to death or sustained inhibition of proliferation of all plaque cells. Normal cells would retain

50% gene function in the presence of inhibitor (either from allele A or allele A'). This therapeutic approach is applicable to the eradication of any clonal proliferation of cells in which allele loss has rendered the cells partially haploid.

- 5     •     LOH has been described in a wide variety of premalignant conditions such as metaplasia and dysplasia of colonic epithelium, breast epithelium, lung epithelium and cervical epithelium. Most studies have focused on metaplastic or dysplastic epithelium adjacent to cancer tissue, and have shown patterns of LOH similar to those in the adjacent malignant  
10     epithelium. Prophylactic ablation of such premalignant tissues could prevent the subsequent development of cancer.
- In benign tumors such as leiomyomas and parathyroidomas, which frequently must be surgically removed, LOH has been well described. As with atherosclerotic plaques, these tumors are frequently multifocal and  
15     therefore the approach of serial inhibition of allele A followed by inhibition of allele A' would alternately abolish essential gene function in one half of all haploid tumor cells and then the other half, leading eventually to death or sustained inhibition of proliferation of all tumor cells.
- 20     •     LOH has been described in endometriosis, a proliferative condition associated with pain and infertility and frequently requiring surgical removal of endometrial tissue growing outside the uterine cavity. As with atherosclerotic plaques, there is only one study published to date and the frequency of LOH is low (15-18%), however the study examined only six  
25     chromosome arms; additional studies may lead to identification of regions of higher frequency LOH
- LOH is apparently the necessary event in the development of cyts in some, and possibly all, forms of autosomal dominant polycystic kidney disease (ADPKD). (There are three forms, with ADPKD1 accounting for about

- 85% of cases and ADPKD2 about 15% of cases.) LOH has been demonstrated by genetic analysis of the cells lining cyst walls in kidneys of ADPKD1 patients: the cells have undergone LOH for markers flanking the ADPKD1 gene. As a result the cyst cells lack functional ADPKD1.
- 5 (Patients with ADPKD inherit one defective copy of an ADPKD gene from their parents.) Only about 20% of cysts were shown to have LOH when studied with a few markers, but this likely reflects, at least to some extent, technical difficulties in obtaining pure populations of cyst cells for analysis. The extent of loss of heterozygosity in cyst cells has not been
- 10 well studied; only several polymorphic markers in the vicinity of the ADPKD1 gene on chromosome 16p were tested in one study (Qian, F., Watnick, T.J., et al. The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. *Cell* 87:979-987, 1996.) Another study found one case of LOH on chromosome 3p, distant
- 15 from the ADPKD gene. Future LOH studies may reveal more extensive LOH in ADPKD. Also, it is worth noting that, unlike malignancy where it is desirable to eradicate all disease cells, eradication of a fraction of the cysts in ADPKD would be expected to have a significant beneficial effect. This is evident from the disparate clinical presentation of ADPKD, with
- 20 varying numbers of cysts being associated with varying degrees of impairment of kidney function.
- Other conditions in which LOH has been demonstrated include hamartomas in tuberous sclerosis patients, odontogenic keratocysts and pterygia (benign lesions of the corneconjunctival limbus).
  - 25 ● Other conditions in which there is evidence of clonal proliferation include inflammatory pannus in arthritic joints, benign prostatic hypertrophy, and hereditary hemorrhagic telangiectasia. (Qian, F. and G.G. Germino. "Mistakes Happen": Somatic Mutation and Disease. *Am. J. Hum. Genet.* 61: 1000-1005, 1997.)

Thus, consistent with the Summary above, it was found that LOH occurs in many non-malignant neoplasias or tumors with subsequent clonal growth of cells which contain only one allelic form in individuals whose normal somatic cells are heterozygous for the particular essential gene. The essential gene can therefore  
5 be inhibited by an allele specific inhibitor, *i.e.*, a variance specific inhibitor. In some conditions, however, multiple, independently arising lesions in an individual are subjected to LOH in a disease or condition, *e.g.*, in the development of atherosclerotic plaques. For that example, in individuals heterozygous for a particular essential gene which undergoes LOH, this results in  
10 some atherosclerotic plaques in which cells have one of the allelic forms of an essential gene, and other plaques in which cells have the alternative form of the gene.

It was determined that such conditions can be treated using allele specific inhibitors despite the presence of both alleles in cells related to the condition.  
15 There are two strategies for such therapy. The first is to serially administer different inhibitors targeted to the different allelic forms of the target gene. This can be accomplished by using inhibitors which target the alternative sequence variants of one sequence variance site. Simultaneous administration of inhibitors of both allelic forms of an essential gene would inhibit the cells which have  
20 undergone LOH at that gene, but would also inhibit the normal heterozygous cells of the individual. This treatment would inhibit essential functions in normal cells as well as cancer cells and have no advantage over the administration of conventional antiproliferative drugs, many of which are inhibitors of known essential functions. In contrast, administration of the first inhibitor targets the  
25 subset of cells which have only the first allelic form of an essential gene. As described for the general strategy, this inhibitor will not significantly affect the growth or survival of the normal heterozygous somatic cells. This first administration is followed by administration of a second inhibitor; the second

125

232/116

inhibitor targets the cells which contain only the second allelic form of the gene, and again does not significantly affect the normal somatic cells. This process of alternating administration can be repeated as needed to achieve a desired therapeutic effect. In some cases many rounds of alternating administrations will  
5 be useful. Similarly, recurring, or even indefinitely continued alternating administrations will provide useful treatment. Likewise, these methods can incorporate the use of inhibitors targeted to specific alleles of a plurality, e.g., 2, 3, 4, or more different target genes.

In certain instances, even though the lesions in non-malignant diseases are not  
10 clonal, there may be systematic loss of one parental chromosome allowing effective therapy with only one variance-specific inhibitor. This would occur, for example, if there were an inherited or early embryonic mutation within a tumor suppressor gene on one parental chromosome, in which case any event which was associated with the elimination of the corresponding normal tumor suppressor  
15 gene on the other parental chromosome would lead to abnormal proliferation. In such cases a variance-specific inhibitor of an essential gene that was closely linked to the normal tumor suppressor gene would preferentially kill cells in the proliferating lesion.

## **VI. Characteristics of allele-specific inhibitors**

20 As indicated above "allele specific inhibitors" or " allele specific anti-neoplastic agents" represent a new approach to tumor therapy because they are lethal or significantly inhibit the growth only of tumor cells. The advantages of this approach include, first, lack of toxicity to the normal cells of the patient resulting in a therapeutic index greater than that of conventional tumor, e.g., cancer  
25 chemotherapy drugs, and second, it is not necessary that the inhibitors be targeted specifically to the tumor cells, as they can be administered systemically. As also described above, usually an allele specific inhibitor is specific for a single



sequence variance of an essential gene, though in some cases the inhibitor utilizes the joint effects of two or more sequence variances on a particular allele.

It is not necessary for the allele specific inhibitor to have absolute specificity. Normal cells expressing equal amounts of two allelic forms of a gene product  
5 encoded by the essential gene will often show a reduction in gene activity when they take up the inhibitors of this invention, but should remain viable due to the activity of the protein encoded by the uninhibited allele. On the other hand, tumor cells expressing only one allele due to LOH, will respond to the inhibitors of this invention which are specifically directed to the remaining allele, with a  
10 greater reduction in gene activity. Growth of tumor cells exposed to the inhibitors of this invention will be inhibited due to the suppression of either the synthesis or the biological activity of the essential gene product.

Also, while a single gene has only two allelic forms in any given individual, the gene can have more than two allelic forms in a human population. Accordingly,  
15 inhibitors can be targeted to any of the alleles in the population. A particular inhibitor will generally be targeted to a subset of the allelic forms; the members of the subset will have a particular sequence variance which provides the specific targeting. In some cases, however, the inhibitor will jointly target two, or possibly more sequence variances.

20 Once two or more alleles are identified for a target essential gene, inhibitors of high specificity for an allele can be designed or identified empirically. Inhibitors that can be used in the present invention will depend on whether allelic variation at a target locus affects the amino acid sequence, the mRNA sequence, or the DNA in intron and promoter regions. If there is variation at the protein level,  
25 then classes of inhibitors would include low molecular weight drugs, oligopeptides and their derivatives, and antibodies, including modified or partial

- antibody fragments or derivatives. For mRNA or DNA sequence variance the main class of inhibitors are complementary oligonucleotides and their derivatives and catalytic RNA molecules such as ribozymes, including modified ribozymes. The generation of inhibitors of this invention can be accomplished by a number of
- 5 methods. The preferred method for the generation of specific inhibitors of the targeted allelic gene product uses computer modeling of both the target protein and the specific inhibitor. Other methods include screening compound libraries or microorganism broths, empirical screening of libraries of peptides displayed on bacteriophage, and various immunological approaches.
- 10 Further, in the treatment of cancer patients, a therapeutic strategy includes using more than one inhibitor of this invention to inhibit more than one target. In this manner, inhibitors directed to different proteins essential to cell growth can be targeted and inhibited simultaneously. The advantage of this approach is to increase the specificity of the inhibition of proliferation of cancer cells, while at
- 15 the same time maintaining a low incidence of side effects.

#### A. Targeted Drug Design.

- Computer-based molecular modeling of target proteins encoded by the various alleles can be used to predict their three-dimensional structures using computer visualization techniques. On the basis of the differences between the three-
- 20 dimensional structure of the alternate allelic forms of the proteins, determinants can be identified which distinguish the allelic forms. Novel low molecular weight inhibitors or oligopeptides can then be designed for selective binding to these determinants and consequent allele-specific inhibition. Descriptions of targeted drug design can be found, for example, in I. Kuntz, "Structure-Based Strategies
- 25 for Drug Design and Discovery," *Science* 257:1078-1082 (1992) and J. Dixon, "Computer-Aided Drug Design: Getting the Best Results," *Trends in Biotechnology* 10:357-363 (1992). Specific applications of the binding of

- molecules to receptors using computer modeling have been described in Piper *et al.*, "Studies Aided by Molecular Graphics of Effects of Structural Modifications on the Binding of Antifolate Inhibitors to Human Dihydrofolate Reductase," *Proc Am. Assoc. Cancer Res. Annual Meeting* 33:412 (1992); Hibert *et al.*, "Receptor 3D-Models and Drug Design," *Therapie* (Paris) 46:445-451 (1991)(serotonin receptor recognition sites). Computer programs that can be used to conduct three-dimensional molecular modeling are described in G. Klopman, "Multicase 1: A Hierarchical Computer Automated Structure Evaluation Program," Quantitative Structure-Activity Relationships, 11:176-184 (1992); Pastor *et al.*, "The Edisdar Programs Rational Drug Series Design," Quantitative Structure-Activity Relationships, 10:350-358 (1991); Bolis *et al.*, "A Machine Learning Approach to Computer-Aided Molecular Design," *J. Computer Aided Molecular Desig*, 5:617-628 (1991); and Lawrence and Davis, "CLIX: A Search Algorithm for Finding Novel Ligands Capable of Binding Proteins of Known Three-Dimensional Structure," *Proteins Structure Functional Genetics* 12:31-41 (1992).

- Low molecular weight inhibitors specific for each allelic protein form can be predicted by molecular modeling and synthesized by standard organic chemistry techniques. Computer modeling can identify oligopeptides which block the activity of the product of the target gene. Techniques for producing the identified oligopeptides are well known and can proceed by organic synthesis of oligopeptides or by genetic engineering techniques. R. Silverman, The Organic Chemistry of Drug Design and Drug Action, Academic Press (1992).

- The inhibitors of this invention can be identified by selecting those compounds that selectively inhibit the growth of cells expressing one allelic form of a gene, but do not inhibit the activity of the A allelic form.

#### B. Small Molecule Inhibitors

129

232/116

Low molecular weight inhibitors can be identified and generated by at least one of the following methods; (1) screening of small organic molecules present in microorganism fermentation broth for allele-specific activity; or (2) screening of compound libraries. Once a compound is identified which exhibits allele specific activity, derivatives of that compound can be obtained or produced in order to obtain compounds having superior properties, such as greater activity, greater specificity, or better administration related properties (*e.g.*, solubility, toxicity, and others).

A small molecule for allele specific targeting, *i.e.*, variance specific targeting, to a polypeptide or protein target will generally have the following characteristics:

- Differential binding affinity for protein domains altered by the amino acid variance *or* uniform binding to the protein with differential effects due to subsequent interactions with variant residues.
- Inhibition of protein function following differential binding. Several mechanisms of inhibition are possible including:
  - competitive inhibition of active sites or critical allosteric sites,
  - allosteric inhibition of protein function,
  - altering compartmentalization or stability, and
  - inhibition of quaternary associations.
- Favorable pharmaceutical properties, such as safety, stability, and kinetics.

In view of the art relating to identification of compounds that interact with particular features of a polypeptide or protein or protein complex, There are clear precedents for developing drugs, *i.e.*, inhibitors, that are variance-specific including drugs that are allosteric inhibitors of protein functions. Several lines of experimental evidence demonstrate that small molecule variance specific

130

232/116

inhibitors can be designed and constructed for particular targets. Specifically:

- Several essential gene targets have been identified that contain variances within domains comprising the active site.
- It is possible to screen for ligands that recognize variant surface features.
- 5 Combinatorial methods using antibodies, peptides, or nucleic acids suggest that specific ligands can be selected for large fractions of the surface of any protein.
- There are many literature reports of single amino acid substitutions, within the active site as well as elsewhere within a protein, altering ligand
- 10 specificity and drug action.
- Allosteric (noncompetitive) inhibition of protein function may be induced by binding ligands to many different surfaces of a protein. Ligands can cause allosteric inhibition by disturbing secondary, tertiary or quaternary (subunit-subunit) interactions of a protein. There is ample evidence that
- 15 such effects can be induced by binding to sequences outside the active site and even in regions that are uninvolved in the normal catalytic or regulatory activity of a protein.

Each of these points is discussed in more detail below.

*Variances located within domains comprising the active site.*

- 20 Crystal structures are available for several of the exemplary targets or for homologous proteins that can allow prediction of tertiary structure. As noted, the protein variance in Replication Protein A occurs within the domain that is involved in binding DNA. The protein variance in CARS occurs within the domain involved in tRNA binding.
- 25 The proximity of the active site to these variances may be exploited by several different strategies:

- Competitive inhibitors can exert variance-specific effects by exhibiting differential affinities for variant active sites, thereby interfering with binding of the substrate or critical allosteric effectors.
- Competitive inhibitors may bind with equal affinity for the active site but exerting different effects on the structure or function of the variant domain.
- Allosteric inhibitors can exert variance-specific effects by binding differentially to variant forms of the active domain and distorting the structure or function of the active site.

10

*Screening for ligands that recognize variant surface features.*

Combinatorial libraries of antibodies, peptides, nucleic acids, or carbohydrates have been used to demonstrate that ligands can be identified that will bind to large fractions of the surface of any protein.

- 15 A library of  $6.5 \times 10^{10}$  antibody-bearing phage was screened for binding to various targets and contained antibodies against all targets tested.

Selex and Aptamer technologies involve selection of random oligonucleotides that bind to specific targets. Reports indicate that ligands with high affinity and specificity can be selected for diverse targets despite the limited chemical

- 20 diversity of the nucleic acid-based ligands.

These studies demonstrate the ability to identify ligands for unique surface features using several different chemistries. Similarly, small molecule protein surface interaction can be screened; two broad approaches for identifying small molecule ligands can be distinguished:

- 25 □ Combinatorial approaches coupled with methods for high-throughput screening provide a similar scope of opportunities as combinatorial methods focused on nucleic acids, peptides, or carbohydrates.

- Rational design or focused combinatorial approaches based on biochemical, biophysical, and structural data about the target protein may be optimal when the crystal structure of the protein is known.. When the crystal structure of the target protein or its homologues are known it will often be possible to model the topology and surface chemistry of the target in detail. These data are useful in optimizing the binding specificity or allosteric inhibitory function of the product through a series of iterative steps once a prototype binding ligand is identified. Structural modeling of the target can be particularly useful in optimizing the variance specificity of a ligand that binds to the target sequence.

*Examples of single amino acid substitutions altering sensitivity to small molecules*

- Many amino acid substitutions have been described in proteins that alter the specificity or function of small-molecule ligands. These substitutions are useful models for variance-specific interactions (*e.g.* interactions that are altered by the amino acid substitutions that distinguish variant forms of a protein.)

There are clear precedents for variance-specific drug effects in humans.

Variance-specific interactions are observed in a wide variety of structurally and functionally heterogeneous proteins. Among these are variances in human proteins including:

- N-acetyl transferase 2 - variances affect acetylation of drugs including caffeine and arylamines;
- CYP2C19 - variances affect the hydroxylation of mephenytoin and related compounds;
- CYP2D6 - variances affect hydroxylation of debrisoquine and related compounds;
- glucose-6-phosphate dehydrogenase - variances account for sensitivity to primaquine and other drugs.

There are numerous examples of variance-specific drug effects in targets for antiviral and antimicrobial drugs. The most extensively characterized are those in HIV Reverse Transcriptase and  $\beta$ -lactamase. These data indicate that many different amino acid substitutions can alter drug effects. Moreover, while amino acid substitutions are classically distinguished as "conservative" or "non-conservative," it is evident from these data that many seemingly "conservative" substitutions can have significant effects. For each of the types of amino acid substitution identified within the exemplary target genes, examples of the same amino acid substitution altering the interaction of small molecule drugs on a target protein is shown in one or more of the model systems.

#### *Sites of allosteric inhibition*

Most drug development focuses on *competitive* inhibitors of protein action rather than noncompetitive, *allosteric* inhibitors. There is no *a priori* advantage to a competitive versus allosteric inhibitor except for the fact that medicinal chemistry often begins with candidate molecules derived from natural substrates or cofactors. There are, in fact, conceptual advantages to allosteric inhibitors since each protein may contain multiple allosteric sites, and allosteric inhibitors may be effective at lower concentrations (*e.g.* those equivalent to the substrate) since there is no need to compete with the substrate for binding.

Detailed crystallographic and other structural studies of a variety of enzymes show that the mechanism of allosteric inhibition commonly involves conformational changes (*e.g.* domain movements) far from the site of contact with the allosteric regulator. These data illustrate the cooperativity of protein structure, demonstrating how a small change in one region of a protein is amplified throughout the structure. Such cooperativity allows small molecules binding to various regions of a protein to have significant structural and



functional effects.

- One way to assess the probability of achieving allosteric effects from a variant sequence is to examine the distribution and nature of mutations that affect drug action in several well-characterized proteins. Another is to examine the
- 5 distribution of epitopes for antibodies that bind to the surface of a protein and inhibit its function. Analyses of these types show that allosteric sites are widely dispersed within proteins and may comprise the majority of the protein's surface.

For example:

*HIV-1 reverse transcriptase (RT)* is a heterodimer with p66 and p51 subunits.

- 10 The p66 subunit is 560 amino acids, and p51 is a 440 amino acid subfragment of p66. The three dimensional structure of HIV-1 RT has been solved by x-ray crystallography. Three HIV-1 RT structures have been published, including complexes with double stranded DNA at 3.0 Å resolution and with the non-nucleoside inhibitors nevirapine (at 3.5Å) and -APA (at 2.8Å).
- 15 Two classes of HIV-1 RT inhibitors have been developed. The first class comprises nucleoside analogues including AZT, ddI and ddC. The second class comprises non-nucleoside analogues belonging to several chemical groups, including TIBO, BHAP, HEPT, -APA, dipyrindiazepinone, pyridinone, and inophyllum derivatives, all of which bind the same hydrophobic pocket in HIV
- 20 RT. Many amino acid substitutions have been described that produce resistance to these drugs. Table 5 shows the location of selected mutations within HIV-1 RT that cause resistance to nucleoside analogues as well as the mechanism of inhibition postulated from physical-chemical experiments and structural data; the list is not comprehensive.

Location and postulated mechanism of amino acid substitutions which confer resistance to nucleoside analog inhibitors. trp266X - multiple substitutions.

Potential resistance mechanism

	Mutation	Location of mutation	Mutation creates resistance to drug(s)	Direct effect on dNTP binding	Indirect effects via interactions with dNTP binding site	Indirect effect by
5	met41leu	a4	AZT		X	
	lys65arg	3- 4	ddC, ddI, 3TC			X
	asp67asn	3- 4 loop	AZT			X
	thr69asp	3- 4 loop	ddC			X
	lys70arg	3- 4 loop	AZT			X
10	leu74val	4	ddI			X
	val75thr		ddI, ddA			
	glu89gly	5a	ddI, ddA			X
	ile135thr	7- 8 loop	ddI		X	
	met184val	9- 10 turn	ddI, ddC	X		X
15	thr215tyr	11a	AZT		X	X
	thr215phe	11a	AZT		X	X
	lys219gln	11b	AZT	X	X	X
	trp266X	-thumb	AZT			

- 20 These data demonstrate that nucleoside analog resistance arises from mutations in multiple domains. Many of the mutations are located far from the dNTP binding sites. These changes inhibit drug function by altering the conformation of the target protein in a manner analogous to those conformational changes that may be induced by an allosteric inhibitor.

Table 5 summarizes the mutations that alter the function of non-nucleoside inhibitor drugs

**Table 5**

Location and postulated mechanism of amino acid substitutions which confer resistance to non-nucleoside analog inhibitors.

Mutation	Mutation location	Effect of mutation	Mutation confers resistance to:
ala98gly	5b- 6 loop	flexibility	Pyridinone L-697661, Nevirapine
leu100ile	5b- 6 loop	-branch	Pyridinone L-697661, Nevirapine, TIBO R82913
lys101glu	5b- 6 loop	charge	Pyridinone L-697661, Pyridinone L-697639,
10 lys103asn	5b- 6 loop	charge loss	Pyridinone L-697661, BHAP U-87201, Nevirapine TIBO R82913
val106ala	6	less bulky	Nevirapine, TIBO R82913
val108ile	6	bulkier	Pyridinone L-697661, Nevirapine
glu138lys	7- 8 loop	charge	TIBO R82913
val179asp	9	charge	Pyridinone L-697661
15 val179glu	9	charge	Pyridinone L-697661
tyr181cys	9	less bulky	Pyridinone L-697661, BHAP U-87201, Nevirapine, TIBO R82913
tyr188cys	10	less bulky	Nevirapine
tyr188his	10	less bulky	TIBO R82913, BHAP U-87201
gly190glu	10	charge	Nevirapine
20 leu228phe	12	bulkier	BHAP U-90152
glu233val	13	charge	BHAP U-87201
pro236leu	13- 14 loop	flexibility	BHAP U-87201
lys238thr	14	charge	BHAP U-87201
trp266X	-thumb		TIBO R82913

It is evident from these examples that the substitutions which inhibit drug functions are distributed across several domains. Different inhibitory mechanisms have been postulated in domains throughout the protein, based on the three-dimensional structure of the protein. Most involve conformational disruption of the protein  
5 secondary and tertiary structure.

*Thyrotropin receptor* Naturally occurring antibodies against the thyrotropin receptor can cause activation of thyroid function (Grave's disease) or inhibition of thyroid function (Hashimoto's disease). The sites within the thyrotropin receptor that are targeted by these natural antibodies have been mapped in detail and have  
10 been tested with monoclonal antibodies. Most of the inhibitory antibodies do not interfere with binding of thyrotropin to its receptor, and thus, are allosteric rather than competitive inhibitors. Several independent classes of inhibitory antibodies have been identified that bind to epitopes within different domains of the receptor. At least one of these epitopes is in a domain that is entirely unimportant for receptor  
15 activity and can be deleted by site-directed mutagenesis without disrupting the function of the receptor. These experiments provide an explicit precedent for achieving allosteric inhibitory effects from ligands that target widely dispersed sequences within the protein.

*Thermus aquaticus DNA polymerase* The inhibitory activity of 24 monoclonal  
20 antibodies to *Thermus aquaticus* DNA polymerase has been investigated. The antibodies recognized 13 non-overlapping epitopes. Antibody binding to eight epitopes was inhibitory. Inhibitory antibodies mapped to several distinct domains, including the 5' nuclease domain, the polymerase domain and the boundary region between the 5' nuclease and polymerase domains. Some antibodies recognized  
25 epitopes overlapping the DNA binding groove of the polymerase. Significantly, the inhibitory antibodies recognized epitopes constituting as much as 50% of the Taq polymerase surface, and the non-inhibitory antibodies a further ~25%.

*β-lactamase* The  $\beta$ -lactamases are a diverse family of enzymes which catalyze the hydrolysis of the  $\beta$ -lactam ring of penicillin and cephalosporin antibiotics.

Interactions of these proteins with various small molecule drugs have been characterized in detail as the pharmaceutical industry has worked to develop

- 5 chemically modified penicillins and cephalosporins to elude inactivation by  $\beta$ -lactamases. In addition, a  $\beta$ -lactamase inhibitor (clavulanic acid) has also been introduced into clinical use.

- As each new drug has been introduced into wide use, mutant  $\beta$ -lactamases have emerged that are resistant to the drug. Over 190  $\beta$ -lactamases have been described  
10 with differential specificity for the various penicillins and cephalosporins. Many of these differ by only a few amino acids. Many different amino acid substitutions at various sites within the protein can change the substrate specificity of the enzyme.

- kat G (Isoniazid resistance)* The *kat G* protein of *M. tuberculosis* encodes a catalase-peroxidase enzyme that is one of two mycobacterial genes frequently  
15 altered in isoniazid resistant strains (the other is *inhA*). There are a wide variety of amino acid substitutions in *katG* associated with drug resistance distributed evenly across the 740 amino acids of the protein. The mechanism by which some of these substitutions inhibit *katG* function can be inferred from the structure of the homologous yeast and *E. coli* enzymes and knowledge of the catalytic function of  
20 the enzyme. For example, insertion of an Ile between positions 125 and 126 affects a conserved interhelical loop near the active site residues; substitutions at amino acid 275 and 315 are likely to affect the ligand access channel; substitutions at amino acid 463 may affect a N-terminal substrate binding site. Other substitutions occur in regions that are not directly related to the functional sites of the protein.

- 25 The examples described above demonstrate that small molecules can discriminate in activity between polypeptides or proteins which have one a single amino acid

difference in sequence, *i.e.*, a single amino acid sequence variance.

The application of small molecule inhibitor identification is specifically discussed in Example 39 below in connection with the methylguanine methyltransferase gene.

5           **C.     Antibody Inhibition.**

Once an essential gene is identified and is determined to exist in two or more allelic forms which encode different proteins, antibodies can be raised against both allelic forms of the protein. The techniques for using a specific protein or an oligopeptide as an antigen to elicit antibodies which specifically recognize epitopes on the  
10 peptide or protein are well known. Preferably monoclonal antibodies (MABs) are used.

In one embodiment, the DNA sequence of the desired allelic form of the target gene can be cloned by insertion into an appropriate expression vector and translated into protein in a prokaryotic or eukaryotic host cell. The protein can be recovered and  
15 used as an antigen to elicit the production of specific antibodies. In another embodiment, the DNA of the desired allelic form of the target gene is amplified by PCR technology and is subsequently translated *in vitro* into protein to be used as the antigen to elicit the production of specific antibodies. A third embodiment is to use the DNA sequence of the alternative alleles as a basis for the generation of synthetic  
20 peptides representing the amino acid sequence of the alleles for use as antigen to elicit the production of specific antibodies.

Antibodies can be generated either by standard monoclonal antibody techniques or generated through recombinant based expression systems. See generally, Abbas, Lichtman, and Pober, Cellular and Molecular Immunology, W.B. Saunders Co.  
25 (1991). The term "antibodies" is meant to include intact antibody molecules of the

IgD isotype as well as antibody fragments or derivatives, such as Fab and F(ab')<sub>2</sub>, which are capable of specifically binding to antigen. The antibodies so produced will preferentially bind only the protein produced in the allelic form which was used as an antigen to create the antibody. If the targeted protein is expressed on the cell surface, the antibody or antibody derivative can be tested as a therapeutic.

Antibody inhibitors are most effective when they are directed against cell surface proteins or receptors. If the essential protein produced by the targeted allele is not a cell surface protein or receptor, the development of antibody inhibitors may also require the use of a special antibody-delivery system to facilitate entry of the antibody into the tumor cells. The plasma membrane that surrounds all cells is designed to limit the entrance of most compounds. Entry is generally restricted to small, non-charged molecules (absence of charge allows them to slip through the fatty membrane) or to those factors that can penetrate the cell using existing, specialized import mechanisms. The introduction into cells of much larger molecules, such as specific antibodies, other proteins, or peptides, requires appropriate delivery systems such as are known in the art. Alternatively, the structure of the variable region of allele specific antibodies can be used as the basis for design of smaller allele specific inhibitory molecules.

#### D. Oligopeptides

Oligopeptides can be demonstrated to have a very high degree of specificity in their interaction with functional polypeptides such as cellular enzymes, receptors or other polypeptides essential for cell viability. Methods for screening peptide sequences which have high specificity for binding to, and functional inhibition of, a specific polypeptide target have been well described previously. Scott, J.K. and Smith G.P., "Searching for Peptide Ligands with an Epitope Library," *Science* 249:386-390 (1990). These methods include the screening of M13 libraries by "phage display" of polypeptide sequences as well as direct screening of peptides or mixtures of synthetic peptides for binding to or inhibition of the target functional polypeptide.

141

232/116

The oligopeptides of this invention can be synthesized chemically or through an appropriate gene expression system. Synthetic peptides can include both naturally occurring amino acids and laboratory synthesized, modified amino acids.

Also provided herein are functional derivatives of a polypeptide or protein. By  
5 "functional derivative" is meant a "chemical derivative," "fragment," "variant,"  
"chimera," or "hybrid" of the polypeptide or protein, which terms are defined  
below. A functional derivative retains at least a portion of the function of the  
protein, for example reactivity with a specific antibody, enzymatic activity or  
binding activity mediated through noncatalytic domains, which permits its utility in  
10 accordance with the present invention.

A "chemical derivative" of the complex contains additional chemical moieties not  
normally a part of the protein. Such moieties may improve the molecule's  
solubility, absorption, biological half life, and the like. The moieties may  
alternatively decrease the toxicity of the molecule, eliminate or attenuate any  
15 undesirable side effect of the molecule, and the like. Moieties capable of mediating  
such effects are disclosed in Remington's Pharmaceutical Sciences (1980).

Procedures for coupling such moieties to a molecule are well known in the art.  
Covalent modifications of the protein or peptides are included within the scope of  
this invention. Such modifications may be introduced into the molecule by reacting  
20 targeted amino acid residues of the peptide with an organic derivatizing agent that is  
capable of reacting with selected side chains or terminal residues, as described  
below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and  
corresponding amines), such as chloroacetic acid or chloroacetamide, to give  
25 carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are  
derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-



alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloro-mercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

- Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

- Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing primary amine containing residues include imidoesters such as methyl picolinimide; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

- Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high  $pK_a$  of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine  $\alpha$ -amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction carbodiimide (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide.

Furthermore, aspartyl and glutamyl residue are converted to asparaginy and  
5 glutaminy residues by reaction with ammonium ions.

Glutaminy and asparaginy residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

- 10 Derivatization with bifunctional agents is useful, for example, for cross-linking component peptides to each other or the complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobi-  
15 functional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl] dithiolpropioimide yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble  
20 matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

- Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino  
25 groups of lysine, arginine, and histidine side chains (Creighton, T.E., Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86

144

232/116

(1983)), acetylation of the Nterminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or  
5 attenuate any undesirable side effect of the protein complex and the like. Moieties capable of mediating such effects are disclosed, for example, in Remington's Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the protein or polypeptide having a length less than the full-length  
10 polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

15 Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide which either lack one or more amino acids or contain additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a naturally occurring polypeptide by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify  
20 codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

A functional derivative of a protein or polypeptide with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the art. For example, the modified components  
25 of the functional derivatives may be produced using site-directed mutagenesis

techniques (as exemplified by Adelman et al., 1983, *DNA* 2:183) wherein nucleotides in the DNA coding the sequence are modified such that a modified coding sequence is modified, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above.

- 5 Alternatively, components of functional derivatives of complexes with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art.

#### **E. Complementary Oligonucleotides and Ribozymes**

- Oligonucleotides or oligonucleotide analogs which interact with complementary  
10 sequences of cellular target DNA or RNA can be synthesized and used to inhibit or control gene expression at the levels of transcription or translation. The oligonucleotides of this invention can be either oligodeoxyribonucleotides or oligoribonucleotides, or derivatives thereof, which are complementary to the allelic forms of the targeted essential gene or they can act enzymatically, such as  
15 ribozymes. Both antisense RNA and DNA can be used in this capacity as chemotherapeutic agents for inhibiting gene transcription or translation. Trojan, J., et al., "Treatment and prevention of rat glioblastoma by immunogenic C6 cells expressing antisense insulin-like growth factor I RNA," *Science* 259:94-97 (1993). Inhibitory complementary oligonucleotides may be used as inhibitors for cancer  
20 therapeutics because of their high specificity and lack of toxicity.

- Included in the scope of the invention are oligoribonucleotides, including antisense RNA and DNA molecules and ribozymes that function to inhibit expression of an essential gene in an allele specific manner. Anti-sense RNA and DNA molecules act to directly block the translation of mRNA by binding to targeted mRNA and  
25 preventing protein translation or directing RNase mediated degradation of the mRNA. With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, e.g., between -10 and +10 regions of the relevant

146

232/116

nucleotide sequence, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific interaction of the ribozyme molecule to complementary target RNA, followed by a  
5 endonucleolytic cleavage. Within the scope of the invention are engineered hammerhead, hairpin, and other motif ribozyme molecules that catalyze sequence specific endonucleolytic cleavage of RNA sequences encoding a gene product essential for cell survival, growth, or vitality.

Specific ribozyme cleavage sites within any potential RNA target can initially be  
10 identified by scanning the target molecule for ribozyme cleavage sites, such as sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide  
15 sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays. *See*, for example, Draper PCT WO 93/23569. For the present invention, the target site will generally include a sequence variance site as described above.

20 Both anti-sense RNA and DNA molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of RNA and DNA molecules. *See*, for example, Draper, *supra*. hereby incorporated by reference herein. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as, for example, solid phase  
25 phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the

antisense or ribozyme RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense or ribozyme cDNA constructs that synthesize antisense or ribozymes RNA

5 constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various modifications to the DNA molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the addition of flanking sequences of ribo- or deoxy- nucleotides

10 to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or methyl phosphonate rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone. Modifications may also be made on the nucleotidic sugar or purine or pyrimidine base, such as 2'-O-alkyl (e.g., 2'-O-methyl), 2'-O-allyl, 2'-amino, or 2'-halo (e.g., 2'-F). A variety of other

15 substitutions are also known in the art and may be used in the present invention. More than one type of nucleotide modification may be used in a single modified oligonucleotide.

A specific application of generating inhibitors which are either complementary oligonucleotides or inhibitory oligopeptides is described in Holzmayer, Pestov, and

20 Roninson, "Isolation of dominant negative mutants and inhibitory antisense RNA sequences by expression selection of random DNA fragments," *Nucleic Acids Research* 20:711-717 (1992). In this study, genetic suppressor elements (GSEs) are identified by random DNA fragmentation and cloning in expression plasmids.

Preferred oligonucleotide inhibitors include oligonucleotide analogues which are

25 resistant to degradation or hydrolysis by nucleases. These analogues include neutral, or nonionic, methylphosphonate analogues, which retain the ability to

interact strongly with complementary nucleic acids. Miller and Ts'O, *Anti-Cancer Drug Des.* 2:11-128 (1987). Further oligonucleotide analogues include those containing a sulfur atom in place of the 3'-oxygen in the phosphate backbone, and oligonucleotides having one or more nucleotides which have modified bases and/or  
5 modified sugars. Particularly useful modifications include phosphorothioate linkages and 2'-modification (*e.g.*, 2'-O-methyl, 2'-F, 2'-amino).

#### F. Gene Therapy

Nucleic acid molecules encoding oligonucleotide or polypeptide inhibitors will also be useful in gene therapy (reviewed in Miller, *Nature* 357:455-460, (1992). Miller  
10 indicates that advances have resulted in practical approaches to human gene therapy that have demonstrated positive initial results. An *in vivo* model of gene therapy for human severe combined immunodeficiency is described in Ferrari, et al., *Science* 251:1363-1366, (1991). The basic science of gene therapy is described in Mulligan, *Science* 260:926-931, (1993).

15 Some methods of delivery that may be used include:

- a. complexation with lipids,
- b. transduction by retroviral vectors,
- c. localization to nuclear compartment utilizing nuclear targeting site found on most nuclear proteins,
- 20 d. transfection of cells *ex vivo* with subsequent reimplantation or administration of the transfected cells,
- e. a DNA transporter system.

A nucleic acid sequence encoding an inhibitor may be administered utilizing an *ex vivo* approach

25 whereby cells are removed from an animal, transduced with the nucleic acid sequence and reimplanted into the animal. The liver can be accessed by an *ex vivo*

approach by removing hepatocytes from an animal, transducing the hepatocytes *in vitro* with the nucleic acid sequence and reimplanting them into the animal (*e.g.*, as described for rabbits by Chowdhury et al, *Science* 254: 1802-1805, 1991, or in humans by Wilson, *Hum. Gene Ther.* 3: 179-222, 1992) incorporated herein by  
5 reference.

Many nonviral techniques for the delivery of a nucleic acid sequence encoding an inhibitor into a cell can be used, including direct naked DNA uptake (*e.g.*, Wolff et al., *Science* 247: 1465-1468, 1990), receptor-mediated DNA uptake, *e.g.*, using DNA coupled to asialoorosomucoid which is taken up by the asialoglycoprotein  
10 receptor in the liver (Wu and Wu, *J. Biol. Chem.* 262: 4429-4432, 1987; Wu et al., *J. Biol. Chem.* 266: 14338-14342, 1991), and liposome-mediated delivery (*e.g.*, Kaneda et al., *Expt. Cell Res.* 173: 56-69, 1987; Kaneda et al., *Science* 243: 375-378, 1989; Zhu et al., *Science* 261: 209-211, 1993). Many of these physical  
15 methods can be combined with one another and with viral techniques; enhancement of receptor-mediated DNA uptake can be effected, for example, by combining its use with adenovirus (Curiel et al., *Proc. Natl. Acad. Sci. USA* 88: 8850-8854, 1991; Cristiano et al., *Proc. Natl. Acad. Sci. USA* 90: 2122-2126, 1993).

In one preferred embodiment, an expression vector containing a sequence encoding a ribozyme or an antisense oligonucleotide is inserted into cells, the cells are grown  
20 *in vitro* and then infused in large numbers into patients.

The gene therapy may involve the use of an adenovirus containing a sequence encoding a ribozyme or an antisense oligonucleotide targeted to a tumor.

Expression vectors derived from viruses such as retroviruses, vaccinia virus, adenovirus, adeno-associated virus, herpes viruses, several RNA viruses, or bovine  
25 papilloma virus, may be used for delivery of nucleotide sequences into the targeted



cell population (e.g., tumor cells). Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors containing coding sequences. See, for example, the techniques described in Maniatis et. al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y. (1989), and  
5 in Ausubel et. al., Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, recombinant nucleic acid molecules encoding protein sequences can be used as naked DNA or in reconstituted system e.g., liposomes or other lipid systems for delivery to target cells (See e.g., Felgner et. al., *Nature* 337:387-8, 1989). Several other methods for  
10 the direct transfer of plasmid DNA into cells exist for use in human gene therapy and involve targeting the DNA to receptors on cells by complexing the plasmid DNA to proteins. See, Miller, supra.

In its simplest form, gene transfer can be performed by simply injecting minute amounts of DNA (e.g., a plasmid vector encoding an inhibitor) into the nucleus of a  
15 cell, through a process of microinjection. Capecchi MR, *Cell* 22:479-88 (1980). The DNA can be part of a formulation which protects the DNA from degradation or prolongs the bioavailability of the DNA, for example by complexing the DNA with a compound such as polyvinylpyrrolidone. Once recombinant genes are introduced into a cell, they can be recognized by the cells normal mechanisms for transcription  
20 and translation, and a gene product will be expressed. Other methods have also been used for introducing DNA into larger numbers of cells. These methods include: transfection, wherein DNA is precipitated with  $\text{CaPO}_4$  and taken into cells by pinocytosis (Chen C. and Okayama H, *Mol. Cell Biol.* 7:2745-52 (1987)); electroporation, wherein cells are exposed to large voltage pulses to introduce holes  
25 into the membrane (Chu G. et al., *Nucleic Acids Res.*, 15:1311-26 (1987)); lipofection/liposome fusion, wherein DNA is packaged into lipophilic vesicles which fuse with a target cell (Felgner PL., et al., *Proc. Natl. Acad. Sci. USA*, 84:7413-7 (1987)); and particle bombardment using DNA bound to small

projectiles (Yang NS. et al., *Proc. Natl. Acad. Sci.* 87:9568-72 (1990)). Another method for introducing DNA into cells is to couple the DNA to chemically modified proteins.

It has also been shown that adenovirus proteins are capable of destabilizing  
5 endosomes and enhancing the uptake of DNA into cells. The admixture of adenovirus to solutions containing DNA complexes, or the binding of DNA to polylysine covalently attached to adenovirus using protein crosslinking agents substantially improves the uptake and expression of the recombinant gene. Curiel  
DT et al., *Am. J. Respir. Cell. Mol. Biol.*, 6:247-52 (1992).

- 10 As used herein "gene transfer" means the process of introducing a foreign nucleic acid molecule into a cell. Gene transfer is commonly performed to enable the expression of a particular product encoded by the gene. The product may include a protein, polypeptide, anti-sense DNA or RNA, or enzymatically active RNA. Gene transfer can be performed in cultured cells or by direct administration into animals.
- 15 Generally gene transfer involves the process of nucleic acid contact with a target cell by non-specific or receptor mediated interactions, uptake of nucleic acid into the cell through the membrane or by endocytosis, and release of nucleic acid into the cytoplasm from the plasma membrane or endosome. Expression may require, in addition, movement of the nucleic acid into the nucleus of the cell and binding to
- 20 appropriate nuclear factors for transcription.

- As used herein "gene therapy" is a form of gene transfer and is included within the definition of gene transfer as used herein and specifically refers to gene transfer to express a therapeutic product from a cell *in vivo* or *in vitro*. Gene transfer can be performed *ex vivo* on cells which are then transplanted into a patient, or can be
- 25 performed by direct administration of the nucleic acid or nucleic acid-protein complex into the patient.

In another preferred embodiment, a vector having nucleic acid sequences encoding an allele specific inhibitor is provided in which the nucleic acid sequence is expressed only in specific tissue. Examples or methods of achieving tissue-specific gene expression are described in International Publication No. WO 93/09236,  
5 published May 13, 1993.

## VII. Utility of allele-specific inhibitors of essential genes

### A. Conditions susceptible to therapy.

The fraction of all cancers could be treated with allele specific inhibitors directed against allele specific essential gene targets is a function of the frequency of the  
10 target allele and the frequency of LOH. The ideal target would be deleted in 100% of all major cancers and would exist in two allelic forms, each with an allele frequency of 0.5 so that half the population would be heterozygous. An inhibitor of one allele of such an ideal target would be a useful agent for 25% of all cancer patients. An inhibitor of the other allele of the same ideal target would be  
15 therapeutic for an additional 25% of all patients, making 50% of all patients treatable. The ideal target has so far not been identified, but we have identified many essential gene sequence variance targets which are deleted in 30-70% of several major cancers, and which are heterozygous in 25-50% of North Americans. Allele specific inhibitors of both alleles of such targets would be expected to  
20 address  $0.4 \times 0.5 = 0.2$  or 20% of the relevant cancer population. The relevant cancer population often includes breast, colon and lung cancer, which sum to ~500,000 new cases per year in the United States. Thus a total available market of 100,000 patients is not unusual, and many targets would be expected to address markets of at least 50,000 patients.

25 The targets of this invention are suitable for treatment of many different cancers, which includes cancers of different types, as well as non-malignant proliferative

153

232/116

disorders, as well as being suitable for use in other applications involving targeting alternative allelic forms of a gene. The classification and nomenclature for a variety of benign and malignant tumors relevant to the present invention is shown in the following table (Table 6-1 from Robbins et al., Pathologic Basis of Disease, 3rd ed. (1984), however, the invention is not limited to these cancers or classifications.

Table 6

	Tissue of Origin	Benign	Malignant
	I. Composed of one parenchymal cell type		<i>Sarcomas</i>
10	A. Tumors of mesenchymal origin		
	(1) Connective tissue and derivatives		
	fibrous tissue	fibroma	fibrosarcoma
15	myxomatous tissue	myxoma	myxocarcoma
	fatty tissue	lipoma	liposarcoma
	cartilage	chondroma	chondrosarcoma
	bone	osteoma	osteosarcoma
			osteogenic sarcoma
20	(2) Endothelial & related tissues		
	blood vessels	hemangioma	angiosarcoma
		capillary	
		cavernous	
		sclerosing	
		hemangioendothelioma	endotheliosarcoma, Kaposi's sarcoma
	lymph vessels	lymphoangioma	lymphangiosarcoma
	synovia		synovioma (synoviosarcoma)
	mesothelium		mesothelioma (mesotheliosarcoma)
25	brain coverings	meningioma	
	glomus	glomus tumor	

		154	232/116
	?endothelial or mesenchymal cells		Ewing's tumor
5	(3) Blood cells & related cells hematopoietic cells		myelogenous leukemia monocytic leukemia malignant lymphomas lymphocytic leukemia plasmacytoma (multiple myeloma)
	lymphoid tissue		histiocytosis X ?histiocytic lymphoma ?Hodgkin's disease
	monocyte-macrophage Langerhans' cells		leiomyosarcoma rhabdomyosarcoma
10	(4) Muscle smooth muscle striated muscle	leiomyoma rhabdomyoma	
	B. Tumors of epithelial origin		<i>Carcinomas</i>
	stratified squamous	squamous cell papilloma	squamous cell or epidermoid carcinoma
15	basal cells of skin or adnexia skin adnexal glands sweat glands sebaceous gland	sweat gland adenoma sebaceous gland adenoma	basal cell carcinoma sweat gland carcinoma sebaceous gland carcinoma
20	epithelial lining glands or ducts -well differentiated group	adenoma papillary adenoma cystadenoma	adenocarcinoma papillary adenocarcinoma cystadenocarcinoma medullary carcinoma undifferentiated carcinoma (simplex) bronchogenic carcinoma bronchial "adenoma"
	poorly differentiated group		melanoma (melanocarcinoma)
	respiratory tract		renal cell carcinoma (hypernephroma)
	neuroectoderm	nevus	
25	renal epithelium	renal tubular adenoma	

		155	232/116
	liver cells	liver cell adenoma	hepatocellular carcinoma
	bile duct	bile duct adenoma	bile duct carcinoma (cholangiocarcinoma)
	urinary tract epithelium (transitional)	transitional cell papilloma	papillary carcinoma transitional cell carcinoma squamous cell carcinoma
5	placental epithelium testicular epithelium (germ cells)	hydatiform mole	choriocarcinoma seminoma embryonal carcinoma
	II. More than one neoplastic cell type---		
10	mixed tumors---usually derived from one germ layer salivary glands	mixed tumor of salivary gland origin (pleiomorphic adenoma)	malignant mixed tumor of salivary gland origin
	renal anlage		Wilms' tumor
15	III. More than one neoplastic cell type derived from more than one germ layer---teratogenous		
20	totipotent cells in gonads or in embryonic rests	teratoma, dermoid cyst	malignant teratoma and teratocarcinoma

Allele specific therapy can be targeted to essential genes which undergo LOH in many different tumor types, including the tumors and tumor types described in the tables above, and in Figure 3.

For the treatment of patients suffering from a tumor using an allele specific inhibitor,

156

232/116

the preferred method of preparation or administration will generally vary depending on the type of inhibitor to be used. Thus, those skilled in the art will understand that administration methods as known in the art will also be appropriate for the inhibitors of this invention.

5           **B.       Pharmaceutical Formulations and Modes of Administration**

The particular compound, antibody, antisense or ribozyme molecule that exhibits allele specific inhibitor activity can be administered to a patient either by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s). In treating a patient exhibiting a disorder of interest, a therapeutically effective amount  
10 of a agent or agents such as these is administered. A therapeutically effective dose refers to that amount of the compound that results in amelioration of one or more symptoms or a prolongation of survival in a patient.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for  
15 determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>. Compounds which exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating  
20 a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized.

For any compound used in the method of the invention, the therapeutically effective  
25 dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating plasma concentration range that

includes the  $IC_{50}$  as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by HPLC.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See *e.g.* Fingl et. al., in The Pharmacological Basis of Therapeutics, 1975, Ch. 1 p.1). It should be noted that the attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, or to organ dysfunctions. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administered dose in the management of the oncogenic disorder of interest will vary with the severity of the condition to be treated and to the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and response of the individual patient. A program comparable to that discussed above may be used in veterinary medicine.

Depending on the specific conditions being treated, such agents may be formulated and administered systemically or locally. Techniques for formulation and administration may be found in Remington's Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990). Suitable routes may include oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections, just to name a few.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For such transmucosal administration,



penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Use of pharmaceutically acceptable carriers to formulate the compounds herein disclosed for the practice of the invention into dosages suitable for systemic  
5 administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular, those formulated as solutions, may be administered parenterally, such as by intravenous injection. The compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for  
10 oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents  
15 may be encapsulated into liposomes, then administered as described above. Liposomes are spherical lipid bilayers with aqueous interiors. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external microenvironment and, because liposomes fuse with cell membranes, are efficiently delivered into the cell  
20 cytoplasm. Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. Determination of the effective amounts is well within  
25 the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. In addition to the active ingredients, these pharmaceutical

compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions. The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable

160

232/116

coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings  
5 for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such  
10 as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added.

Factors specific for the delivery of antisense and ribozyme nucleic acids are known in the art, for example as discussed in Couture et al., WO 94/02595, which is hereby  
15 incorporated herein by reference. This reference also describes the synthesis of nucleic acid molecules having a variety of 2' modified nucleotides.

The references cited herein are incorporated by reference to the same extent as if each had been individually incorporated by reference. The invention is illustrated further by the following examples, which are not to be taken as limiting in any way. The  
20 examples, individually, and together, further demonstrate that one skilled in the art would be able to practice each of the steps in developing useful pharmaceutical products as described in the invention. Generally, the development of such a product involves the following steps:

1. Select candidate target gene essential for cell survival or proliferation.
- 25 2. Determine chromosome location and LOH frequency.
3. Identify common variance in the normal population.

4. Demonstrate antiproliferative effects from inhibition of candidate gene.
5. Design variance-specific inhibitor.
6. Achieve variance-specific antiproliferative effects in cancer cells.

## **EXAMPLES**

### **5 Example 1. Genes required for Cell Proliferation**

Many genes are involved in the process of cell proliferation and are potential targets for anti-proliferative drugs in this invention. Dividing cells progress through a repeating cycle of four stages, each of which is critical to the proliferation process. During the first phase, G1, cells ready the proteins they need to replicate their DNA, which occurs during S phase. Following S phase, cells enter G2, in which they prepare to divide into two daughter cells, each of which will contain the DNA content of the original cell. The final stage of the cell cycle is M phase, in which cells undergo mitosis. During mitosis, the cell nucleus disappears and the two sets of replicated chromosomes are separated to opposite sides of the cell. The cell then divides into two cells, the nucleus reforms in each new cell, and the cycle begins again. Cell proliferation is exceedingly complex and requires the precise coordination of many processes, including DNA synthesis, chromosome condensation and separation, and cell fission. In eukaryotic cells such as yeast, many of the proteins involved in cell division are encoded by essential genes, including those contributing to the duplication of the nucleus and the functions of microtubules, spindle pole bodies the centromere and the kinetochore.

A number of proteins are essential for cell proliferation. Proteins that are critical to this process can be divided into two classes: (i) proteins that regulate cell division; (ii) proteins that form structures involved in cell division. Proteins that regulate cell division include, but are not limited to, proteins involved in the regulation of particular

steps in the division process, such as nuclear breakdown and the transition between the different stages of mitosis, as well as proteins regulating the initiation of mitosis, such as the cyclins, cyclin-dependent kinases (CDKs), and the kinases and phosphatases that regulate CDKs. Cyclin B, the cyclin-dependent kinase *cdc2*, and the *cdc25C* phosphatase are examples of proteins that regulate the initiation of mitosis. Deletion of yeast homologs of these genes is lethal, verifying their critical role in regulating the entry into mitosis. (It has been established that many human genes which encode proteins involved in highly conserved cellular processes can substitute for their yeast counterparts, and vice versa. For example such conservation has been demonstrated for components of the transcriptional apparatus, as well as components of the translational apparatus.)

Proteins that form structures involved in cell division include, but are not limited to, those involved in the processes of chromosome condensation and separation. Examples are tubulin and kinesin, which participate in the separation of chromosomes, and KIAA0165 and CDC37, involved in the spindle pole. Deletion of the yeast homolog of CDC37 is lethal.

Inhibiting the ability of a cell to divide induces, by definition, a cytostatic response, often followed by cell death. Colchicine and nocodazole are examples of drugs that inhibit microtubule function *in vitro*, thereby preventing chromosome separation and leading to cell cycle arrest during mitosis. Vinblastine and vincristine, which also inhibit microtubule function and therefore cell proliferation, have been used widely in the treatment of cancer.

Examples of genes that are involved in the process of cell proliferation, and are thus essential for cell survival or proliferation are shown in the accompanying table. Each of these genes has been disrupted in *Saccharomyces cerevisiae* and the mutant yeast shown to be nonviable.

**Table: Genes Essential for Cell Proliferation in Yeast**

Gene Name	Function of Gene Product
APC1	Component of the anaphase promoting complex.
CAK1	cdk activating kinase, activates cdc28p
CBF2, CBF3B, CSE1 CBF5, CTF13, SKP1	Essential constituents of the kinetochore protein complex Cbf3 (subunits a-d), a structural component of centromeres to which microtubules attach.
CDC14	Protein tyrosine phosphatase that performs a function late in the cell cycle.
CDC15	Essential for late nuclear division
CDC16, CDC23, CDC27	Part of anaphase promoting complex, required for Clb2p degradation and metaphase-anaphase transition.
CDC28	Essential for mitosis
CDC31	Calcium binding protein of spindle pole body (SPB), involved in SPB duplication
CDC37	Required for spindle pole duplication and passage through START.
CDC5	Protein kinase required for exit from mitosis, and operation of mitotic spindle.
CKS1	Associated with cdc28p kinase
CRM1	Chromosome region maintenance protein.
CSE1	Probable kinetochore protein, interacts with centromeric element CDEII.
CSE4	Required for chromosome segregation.
DBF4	Regulatory subunit for cdc7p protein kinase, required for G1/S transition.
DIS3	Involved in mitotic control.
DNA43	Required for S-phase initiation or completion.
DPB11	Involved in DNA replication and an S-phase checkpoint.
ESP1, KAR1	Required for regulation of spindle body pole duplication.
IPL1	Protein kinase involved in chromosome segregation.
KRR1	Essential for cell division.
MEC1	Checkpoint protein required for mitotic growth, DNA repair and recombination.
MIF2	Centromere protein required for chromosome segregation and spindle integrity

164

232/116

MOB1	Required for normal cell cycle progression
MPS1	Protein kinase involved in spindle body pole duplication; also mitotic checkpoint
NDC1	Required for spindle body pole duplication; nuclear envelope component
NNFI	Nuclear envelope protein required for nuclear migration during mitosis.
NRK1	Protein kinase that interacts with cdc31p
NUF2	Component of spindle body pole required for nuclear division.
RFT1	Involved in nuclear division.
SMC1, SMC2, SMC3	Coiled coil proteins involved in chromosome condensation and segregation; required for nuclear division.
SPC42, SPC97, SPC98, SPI6	Components of spindle pole body. The latter 3 interact with microtubules, gamma tubulin & stu2p, respectively.
SPK1	Protein kinase with a checkpoint function in S and G2
STU1	Required for mitotic spindle assembly.
TEM1	Involved in termination of M-phase.

It will be evident to one skilled in the art that many genes that express essential metabolic and homeostatic functions of the cell will also be essential for cell proliferation.

**Example 2. Genes required to maintain inorganic ions at levels compatible with cell growth or survival.**

*Inorganic Ions are Essential for Cellular Life*

Inorganic ions are required for virtually all cellular processes: they are important for maintenance of cell shape and osmolality; they are prosthetic groups of a wide variety of enzymes; they are required for ATP production coupled to ion diffusion; they mediate signal transduction both from intracellular and extracellular signals. Hence maintenance of inorganic ions at physiological concentrations is essential for cell

165

232/116

proliferation and cell survival. The importance of maintaining physiological ion concentrations is further demonstrated by the observation that deviation from normal levels leads to cytostatic or cytotoxic effects, as demonstrated by the effects of selectively poisoning ion channels or placing cells in hypotonic or hypertonic extracellular fluid.

*Inorganic Ions Must be Transported Across Membranes*

Maintenance of ion concentrations at optimal concentrations within cells is complicated by the presence of membranes which, because of their hydrophobic interior, form a highly impermeable barrier to most polar molecules, including inorganic ions. Important cell membranes include the plasma membrane as well as the nuclear membrane, mitochondrial membranes, the endoplasmic reticulum and Golgi apparatus, lysosomes and vesicles of various types, all of which are essential for cell proliferation or survival. Therefore maintaining the concentration of essential polar molecules, including both organic and inorganic ions, at levels compatible with cell growth or survival requires specialized mechanisms for moving such ions across the plasma membrane and the various intracellular membrane bound compartments.

Vital components of the apparatus for maintaining ion concentrations at levels essential for cell survival include regulatory molecules that sense the concentration of ions in different cellular compartments and produce signals to increase or decrease the concentration of said ions to levels compatible with cell survival; proteins that actively or passively transport ions across membranes; and proteins that modify ions so they can be transported across membranes.

Membrane transport proteins can be divided into several categories depending on whether they require energy (provided either by ATP hydrolysis or by co-transport of ions such as sodium or protons down their electrochemical gradients), produce energy



(ATP synthetases, which are usually coupled to proton diffusion) or are energy neutral. Other categories of transporters include those that transport one or more solutes (one or more of which may be ions), gated vs. non-gated - i.e. open only transiently (ligand gated and voltage gated channels) or open continuously, allowing ions to move down their concentration and electrochemical gradients. Specific types of essential membrane transporters include uniports, which simply transport one solute from one side of the membrane to the other, and cotransports, in which the transport of one solute is dependent on the simultaneous or sequential transport of a second solute in the same direction (symport) or in the opposite direction (antiport).

Other inorganic ions, such as iron, are transported bound to carrier proteins (transferrin in the case of iron). Transport of the iron carrier protein involves a complex cycle that begins with binding of iron to transferrin, binding of the iron-transferrin complex to transferrin receptor, formation of coated pits, endocytosis of the transferrin-iron complex via the coated pits, release of iron from transferrin in endosomes upon acidification to pH 5, and then recycling of the transferrin receptor-apotransferrin complex to the surface of the cell where, at neutral pH, the apotransferrin is released from transferrin receptor into the extracellular fluid to bind more iron and participate in another cycle. Thus in the case of transferrin-mediated iron transfer there are a variety of specialized proteins which must interact in a coordinated manner for transport to occur effectively.

Some of the specific inorganic ions which must be transported across the both the plasma membrane and intracellular membranes are sodium, potassium, chloride, calcium, hydrogen, magnesium, manganese, phosphate, selenium, molybdenum, iron, copper, zinc, fluorine, iodine, chromium, silicon, tin and arsenic. Specific transporters have been identified for many of these solutes including sodium, potassium, chloride, protons, copper and iron among others.

Regulation of ion concentrations at appropriate levels is often an energy-dependent process; intracellular and extracellular concentrations may differ by 10 fold or more (see Table).

**Ion Concentrations Inside and Outside a  
Typical Mammalian Cell**

<b>Ion</b>	<b>Intracellular concentration (mM)</b>	<b>Extracellular concentration (mM)</b>
<b>Cations</b>		
Na <sup>+</sup>	5-15	145
K <sup>+</sup>	140	5
Mg <sup>++</sup>	30	1-2
Ca <sup>++</sup>	1-2	2.5-5
<b>Anions</b>		
Cl <sup>-</sup>	4	110

*Inhibitors of Ion Transporting Proteins are Cytostatic or Cytotoxic*

Blocking import of essential cell nutrients, including inorganic ions, prevents cell growth and can lead to cell death. A well studied example is blockade of iron transport by inhibition of transferrin receptor. Dividing cells require iron, and transferrin receptor-mediated uptake of iron-transferrin complexes is the principal route for iron acquisition. Iron uptake requires multiple steps, including receptor binding, endocytosis via coated pits, acidification of endosomes and consequent release of iron from transferrin, followed by recycling of transferrin receptor-apotransferrin to the cell surface for another round of binding. Each step requires the coordinated function of a variety of proteins. Anti-transferrin receptor antibodies arrest cell growth by blocking iron uptake; antitumor effects have been demonstrated *in vitro* and *in vivo* with such antibodies.

Ion pumps are another class of proteins for which cytotoxic inhibitors have been

168

232/116

identified. All animal cells contain a  $\text{Na}^+$ ,  $\text{K}^+$  pump which operates as an antiport, actively pumping  $\text{Na}^+$  out of the cell and  $\text{K}^+$  in against their concentration gradients. In coupling the hydrolysis of ATP to the active transport of 3  $\text{Na}^+$  out and 2  $\text{K}^+$  into the cell the pump is electrogenic. The electrochemical gradients generated and maintained by the  $\text{Na}^+$ ,  $\text{K}^+$  pump are essential for regulation of cell volume and for the secondary, sodium-coupled active transport of a variety of organic and inorganic molecules including glucose, amino acids and  $\text{Ca}^{++}$ . Hence the sodium potassium pump plays an essential role in cellular physiology. More than one third of a typical animal cells energy requirement is expended in fueling this pump. (Alberts et al. Molecular Biology of the Cell, Garland Publishing, New York, 1983, p.291.) Ouabain is an inhibitor of the  $\text{Na}^+$ ,  $\text{K}^+$  ATPase. It binds to the catalytic alpha 1 subunit of sodium potassium ATPase and is a potent cytotoxic drug. Cells treated with ouabain swell and eventually burst as they are unable to maintain a balance of osmotic forces because they can no longer pump out  $\text{Na}^+$ . See Example 11 for a more detailed description of the essential properties of the  $\text{Na}^+$ ,  $\text{K}^+$  ATPase. Amiloride is another cytotoxic drug; it blocks the sodium-proton antiporter. Thus inhibition of proteins essential for maintaining physiological levels of inorganic ions is toxic to cells.

*Ion Transporting Proteins are Evolutionarily Conserved and Essential in Other Species*

Many of the proteins required to maintain inorganic ions at physiologic levels are widely conserved in eukaryotes, reflecting an ancient and vital role. A number of gene disruption experiments in non-human cells demonstrate the importance of ion transporting proteins for cell growth and survival. For example in the yeast *Saccharomyces Cerevisiae* the gene encoding CDC1 protein, involved in maintaining ion homeostasis, has been disrupted resulting in non-viable yeast. Another essential yeast gene is PMA1, which encodes a  $\text{H}^+$  transporting P-type ATPase of the plasma membrane; activity of the encoded protein is rate limiting for growth at low pH.

As a result of the essential functions provided by proteins required for maintenance of inorganic ions at levels required for cell growth or survival, those genes which undergo LOH in a neoplastic disorder and which have sequence variants (nucleic acid or amino acid sequences) in a population as described above, are appropriate potential targets for allele specific inhibition, and thus can be used in the methods for identifying allele specific inhibitors and in other aspects of this invention. The provision of the exemplary ion transport genes, including sodium-potassium ATPase alpha subunit as well as the other genes listed in the Target Genes Table, indicates that other genes within this category or related subcategories will also be appropriate potential targets. Such a gene can be identified as an essential gene by reference to the art, or by the essential gene identification methods known in the art, examples of which are referenced herein. The LOH and sequence variance characteristics can then be readily determined by the described methods, thereby demonstrating that the gene is an appropriate potential target gene for allele specific inhibition.

**Example 3. Genes required to maintain organic compounds at levels compatible with cell growth or survival.**

*Organic Compounds are Essential for Cellular Life*

Organic compounds include the amino acids, carbohydrates, lipids, nucleosides and nucleotides, ions such as bicarbonate, vitamins such as ascorbic acid, pantothenic acid, riboflavin, nicotinamide, thiamine, vitamin B6, vitamin B12, and folate, essential nutrients such as linoleic acid and a wide variety of metabolic intermediates. Organic compounds are required for virtually all vital cellular processes: they are the building blocks of all cellular macromolecules including larger organic compounds such as proteins, starches, polynucleotides and complex lipids as well as glycolipids,

glycoproteins, lipoproteins, etc.; they are constituents of all cell structural molecules including proteins and membranes; they constitute all the metabolic intermediates in such vital cell processes as glycolysis, the Krebs cycle, oxidative phosphorylation, gluconeogenesis, the urea cycle, nucleotide biosynthesis, amino acid biosynthesis, etc.

5 Maintaining organic compounds at levels compatible with cell growth or survival constitutes a large fraction of the work of the cell. Deviation from normal levels of organic compounds will generally have cytotoxic or cytostatic effects on cells (if the appropriate homeostatic cellular machinery for maintaining organic compounds at levels compatible with cell growth or survival is not operating to bring levels back to  
10 normal), as demonstrated by the effects of preventing transport of organic ions such as essential amino acids, vitamins or ions such as bicarbonate or blocking such processes as glycolysis or amino acid biosynthesis or transport of proteins into mitochondria, or required post-translational processing of proteins, lipids or carbohydrates.

15 *Maintaining Organic Compounds at Levels Compatible with Cell Growth or Survival Requires Membrane Transport, Biosynthesis, Energy Extraction, Energy Production, Degradation and Excretion Pathways*

Maintenance of organic compounds at optimal concentrations within cells is  
20 complicated by the presence of membranes which, because of their hydrophobic interior, form a highly impermeable barrier to most polar or charged molecules or molecules over 100 Daltons, including many organic compounds. Important cell membranes include the plasma membrane as well as the nuclear membrane, mitochondrial membranes, the endoplasmic reticulum and Golgi apparatus, lysosomes  
25 and vesicles of various types, all of which are essential for cell proliferation or survival. Therefore maintaining the concentration of essential organic compounds at levels compatible with cell growth or survival requires specialized mechanisms for moving such compounds across the plasma membrane and the various intracellular membrane bound compartments.

Vital components of the apparatus for maintaining organic compounds concentrations at levels essential for cell survival include regulatory molecules that sense the concentration of ions in different cellular compartments and produce signals to increase or decrease the concentration of said compounds to levels compatible with cell survival; proteins that actively or passively transport organic compounds across membranes; and proteins that modify or bind to organic compounds so they can be transported across membranes.

Some of the specific inorganic ions which must be transported across the both the plasma membrane and intracellular membranes are sodium, potassium, chloride, calcium, hydrogen, magnesium, manganese, phosphate, selenium, molybdenum, iron, copper, zinc, fluorine, iodine, chromium, silicon, tin and arsenic. Specific transporters have been identified for many of these solutes including sodium, potassium, chloride, protons, copper and iron among others.

The number of essential membrane proteins is not known. A crude estimate can be derived by adding up the proteins which perform essential functions enumerated above. There are many presently known organic compounds which must be transported across the cell membrane, including small molecules such as essential amino acids, lipids, sugars, the vitamins pantothenic acid, folic acid, riboflavin, nicotinamide, thiamine, vitamin B<sub>6</sub>, vitamin B<sub>12</sub> and ascorbic acid as well as larger molecules such as proteins. (It is important to note that some essential functions are performed by families of transporters with overlapping tissue expression. In such cases it may be that no one protein is essential despite the fact that the protein family collectively carries out an essential cell function. Conversely, there are likely to be a number of essential membrane proteins not yet identified.)

*Examples of Genes Essential to Maintain Organic Compounds at Levels Compatible with Cell Growth or Survival, From Yeast*

The yeast *Saccharomyces Cerevisiae* is a eukaryote which shares many genes in common with humans. Approximately 70% of the essential genes in yeast have human homologs. Many human genes can be exchanged with their yeast counterparts with minimal effects on growth in yeast or human cells. The study of essential genes in yeast is much further advanced than in mammalian systems: over half of the ~6,000 genes of *Saccharomyces Cerevisiae* have been disrupted and the phenotype of the resulting strains tested on minimal growth media. Over 20% of disrupted yeast genes are essential, and a significant fraction of their human counterparts are likely to be essential for cell survival. Among the yeast genes disrupted are a variety of genes that encode proteins required to maintain organic compounds at levels compatible with cell growth or survival. Many of these genes are essential for cell survival. Many of the disrupted essential yeast proteins have closely related human homologs, and it is likely that the human homologs are also essential. Specific examples of yeast genes that are essential are listed below. (This is a partial list; see the web site [\\_\\_\\_\\_\\_.proteome.com](http://proteome.com) for an up to date list.)

The yeast ACC1 gene encodes acetyl co-A carboxylase and, like the human enzyme, is the first and rate limiting step in fatty acid biosynthesis.

The yeast DYS1 gene encodes deoxyhypusine synthase which catalyzes the first step in biosynthesis of the polyamine deoxyhypusine.

The yeast FBA gene encodes fructose-bisphosphate aldolase II, the sixth step in glycolysis, while the essential yeast genes GND1, ENO2, GPM1 and PYK1 encode 6-phosphogluconate dehydrogenase, enolase 2, phosphoglycerate mutase and pyruvate kinase (the last step of glycolysis).

The yeast ERG10 gene encodes acetyl-CoA-acetyltransferase, the first step in the mevalonate/sterol pathway. The essential ERG1 gene encodes squalene

monooxygenase, an later enzyme of the sterol biosynthesis pathway. ERG7, ERG8, ERG9, ERG11, ERG20, ERG24 and ERG25 encode enzymes on the same or related pathways.

5 The yeast ALG1 and ALG2 genes encode mannosyltransferases required for N-glycosylation, and the ALG7, DPM1 and NMT1 genes encode transferases for UDP-N-acetyl-glucosamine-1-P, mannose and myristate, respectively. RAM2 encodes a protein that is a subunit of both farnesyltransferases and (with BET2) geranylgeranyltransferases.

10

The yeast LCB1 gene encodes serine C-palmitoyltransferase which catalyzes the first step in the pathway for synthesis of the long chain base component of shingolipids, while the yeast AUR1 gene encodes a phosphoinositol transferase also essential for shingolipid synthesis.

15

The yeast PRO1 and PRO2 genes encode the three enzymes of proline biosynthesis. THR1 catalyzes the first step of threonine biosynthesis.

20

**Example 4. Genes required to maintain cellular proteins at levels compatible with cell growth or survival.**

25

Proteins carry out a host of essential enzymatic and structural functions required for cell proliferation and cell survival. Consequently, complete inhibition of protein synthesis is eventually lethal to all cells. The requirement of dividing cells (including cancer cells) for high level protein synthesis makes them more sensitive than quiescent cells to the cytostatic and cytotoxic effects of protein synthesis inhibitors. Because the basic scheme of protein synthesis remains the same in all living organisms there are many attractive schemes for screening human targets in heterologous organisms.



*Polypeptide Synthesis Occurs in Several Steps and Requires Over 100 Proteins*

The machinery of polypeptide synthesis includes:

5 Aminoacyl tRNA synthetases, which covalently couple amino acids to their cognate tRNAs. Eukaryotic cells have two sets of tRNA synthetases, one for aminoacylation of cytoplasmic tRNAs and one for aminoacylation of mitochondrial tRNAs. Both types of tRNA synthetases are encoded in the nuclear genome.

10 Ribosomes, which translate mRNA into protein and integrate the action of the other components of the polypeptide polymerization machinery.

Initiation factors, which mediate the steps before the first peptide bond is formed, including formation of an initiation complex consisting of a ribosome, an mRNA and the first aminoacyl tRNA. Initiation is generally the rate limiting step in polypeptide synthesis.

15 Elongation factors, which function in all the reactions between synthesis of the first peptide bond and addition of the last amino acid.

Termination factors, which perform the reactions required to release completed polypeptide chains from ribosomes.

20 Polypeptide chaperonins and other folding factors such as isomerases, which are necessary for the proper folding (and hence function) of proteins.

Polypeptide degradation machinery, including the ubiquitin system for tagging proteins for degradation and the proteasome and its constituents for cleaving proteins targeted for degradation. As cells grow and respond to changing circumstances there is a continual need to new protein synthesis. However,  
25 without some mechanism for eliminating existing unneeded or damaged proteins cells are not able to survive or proliferate.

There are approximately 20 cytoplasmic and 20 mitochondrial tRNA synthetases, approximately 80 ribosomal proteins, and over 20 protein constituents of initiation

175

232/116

factors, elongation factors and termination factors. The available evidence suggests that virtually all of these proteins are encoded by single copy genes. Thus at least 150 genes and their encoded proteins are potential candidates for allele specific targeting. (Conversely, the RNA constituents of the translational apparatus - transfer RNAs and ribosomal RNAs - are encoded by multicopy genes and do not constitute targets for allele specific targeting).

*Inhibitors Have Been Identified for Most Steps of Peptide polymerization and processing*

Well over 100 protein synthesis inhibitors with a wide variety of structures and mechanisms of action have been characterized in both prokaryotes and eukaryotes. Specific inhibitors have been identified for each step of translation described above. See Table from Vasquez (ref. 1) for a summary of translation inhibitors.

Inhibition of aminoacyl tRNA synthetases has been accomplished by at least three different mechanisms: amino acid analogs such as borrelidin and histidinol result in arrest of cell division by competing with natural amino acids for aminoacylation by tRNA synthetases. Inhibition of prokaryotic cell growth has also been demonstrated with RNA minihelices which mimic the acceptor stems of tRNAs. The minihelices compete with authentic tRNAs for aminoacylation by cognate tRNA synthetases. A third class of synthetase inhibitor is represented by pseudomonic acid A, a species specific inhibitor of gram positive isoleucyl tRNA synthetase produced by a gram negative organism. Pseudomonic acid A does not mimic amino acids or tRNAs, but binds to isoleucyl tRNA synthetase to inhibit peptide polymerization and processing.

Peptide polymerization and processing inhibitors that act on ribosomes include agents which bind the protein components and agents which bind or cleave the RNA components of ribosomes. An example of the former is the small

molecule drug emetine, which binds to ribosomal protein S14 and inhibits peptide polymerization and processing.

*Peptide polymerization and processing Inhibitors are Cytostatic or Cytotoxic Drugs*

5

Some of the most potent cytotoxic agents known are protein synthesis inhibitors. For example, a single molecule of ricin or diphtheria toxin is sufficient to kill a cell.

10

The largest class of protein synthesis inhibitors act on the elongation step of translation, with many inhibitors known for both prokaryotes and eukaryotes. Among the best studied prokaryotic elongation inhibitors are molecules belonging to the major antibiotic groups: the tetracyclines, streptomycin and other aminoglycosides, lincomycin and related compounds, erythromycin and related macrolide antibiotics and puromycin. Among the better characterized eukaryotic elongation inhibitors are toxins

15

such as ricin and diphtheria toxin.

*Cancer Chemotherapy by Inhibition of Peptide polymerization and processing*

20

The best studied chemotherapeutic agent that acts solely by inhibiting protein synthesis is the enzyme L-asparaginase, used frequently in the treatment of acute lymphoblastic leukemia and occasionally in the treatment of other cancers. The therapeutic effect of L-asparaginase treatment is hydrolysis of serum L-asparagine to L-aspartate, with a rapidly ensuing drop in serum asparagine levels. While asparagine is not an essential amino acid, leukemia cells generally do not express asparagine synthase and are therefore reliant on importation of asparagine from serum via amino acid transporters in the plasma membrane. The effect of sudden asparagine starvation on rapidly dividing leukemia cells is to induce apoptotic death. Subsequent retreatment with L-asparaginase is generally not as effective as the initial treatment because the leukemia cells which survived the initial treatment have had time to induce expression of

25

asparagine synthase and are no longer dependent on external asparagine.

*Examples of Genes Essential to Maintain Cellular Proteins at Levels Compatible with Cell Growth or Survival, From Yeast*

5

10

15

20

The yeast *Saccharomyces Cerevisiae* is a eukaryote which shares many genes in common with humans. Approximately 70% of the essential genes in yeast have human homologs. Many human genes can be exchanged with their yeast counterparts with minimal effects on growth in yeast or human cells. The study of essential genes in yeast is much further advanced than in mammalian systems: over half of the ~6,000 genes of *Saccharomyces Cerevisiae* have been disrupted and the phenotype of the resulting strains tested on minimal growth media. Over 20% of disrupted yeast genes are essential, and a significant fraction of their human counterparts are likely to be essential for cell survival. Among the yeast genes disrupted are a variety of genes that encode proteins required to maintain proteins at levels compatible with cell growth or survival. Many of these genes are essential for cell survival. Many of the disrupted essential yeast genes have closely related human homologs, and it is likely that the human homologs are also essential. Specific examples of yeast genes that are essential are listed below. All of these genes have human homologs. (This is a partial list because the *Saccharomyces* gene disruption project is only halfway done; see the web site <http://quest7.proteome.com> for an up to date list.)

25

GRC5, NHP2, NIP1, RPL1, RPL25, RPL27, RPL32, RPL35, RPL7, and URP2 are yeast ribosomal proteins that have been disrupted and found to be essential.

CDC33, GCD1, GCD10, GCD11, GCD2, GCD6, GCD7, PRT1, SIS1, SUI1, SUI2, SUI3, TIF11, TIF34, and TIF5 are essential translation factors, mostly translation initiation factors that initiate translation at ATG.

EFB1 and YEF3 are translation elongation factors that have been disrupted and found essential.

SUP35 and SUP45 are essential translation termination factors.

ALA1, HTS1, DED81, THS1, VAS1, WRS1 and KRS1 are essential yeast cytoplasmic tRNA synthetases.

#### *References*

1. Vazquez, D. (1979) Molecular Biology and Biophysics, vol. 30. Inhibitors of Protein Synthesis. Springer-Verlag, Berlin.
2. Lim-Sylianco, C.Y. (1990) Toxins that alter the expression of genetic information: genotoxins and inhibitors of RNA or protein synthesis, pp. 338-421 in Shier, W.T. and D. Mebs, eds., Handbook of Toxinology, Marcel Dekker, New York.

#### **Example 5. Genes required to maintain cellular nucleic acids at levels compatible with cell growth or survival.**

Cellular nucleic acids including deoxyribonucleic acids and ribonucleic acids are essential elements for cell survival and proliferation. Many different genes are involved in maintaining these constituents at levels required for cell growth and proliferation including genes encoding enzymes for nucleotide synthesis, nucleotide degradation and salvage, polymerization of DNA (replication), polymerization of RNA (transcription), modifications of DNA including methylation, modifications of RNA including polyadenylation and capping, and processing of DNA and RNA. Many of these genes and their gene products are targets for conventional antiproliferative drugs.

*RNA and DNA precursor Biosynthesis is Essential for Cell Proliferation*

Nucleotides, the building blocks for both RNA and DNA, are essential for cell survival. Eukaryotic cells have several pathways for the production of nucleotides: de novo purine and pyrimidine biosynthesis, salvage pathways and membrane transport.

5

*Over 50 Proteins Participate in RNA and DNA precursor Biosynthesis*

The principal enzyme groups involved in RNA and DNA precursor biosynthesis are the 14 enzymes of de novo purine biosynthesis, 5 enzymes of de novo pyrimidine biosynthesis (encoded in two polypeptides) and the enzymes of the nucleotide salvage pathways, which number at least 10.

10

*Inhibitors of RNA and DNA precursor Biosynthesis are Cytostatic or Cytotoxic Drugs Useful in Cancer Chemotherapy* Many of the most clinically effective antineoplastic agents block steps in RNA and DNA precursor biosynthesis. Examples include agents which block enzymes of de novo purine and pyrimidine biosynthesis or interfere with salvage pathways. For example, hydroxyurea blocks production of deoxyribonucleotides by ribonucleotide diphosphate reductase.

15

*Purine Biosynthesis is essential for cell proliferation*

Pharmacologic inhibitors of purine biosynthesis are cytotoxic. These include drugs like azaserine and 6-diazo-5-oxo-L-norleucine (DON), glutamine analogs which inhibit three steps in purine synthesis, the most important being inhibition of the enzyme formylglycinamide ribonucleotide amidotransferase. 8-azaguanine and mycophenolic acid interfere with guanylate biosynthesis. (See Komberg, A., DNA Replication, W.H. Freeman and Company, San Francisco, 1980, for a review of drugs that inhibit purine and pyrimidine biosynthesis.) There is also evidence of the essentiality of purine biosynthesis from yeast. For example, the *saccharomyces cerevisiae* PUR5 gene encodes inosine 5'-monophosphate dehydrogenase, which converts inosine 5'-phosphate and NAD to xanthosine 5'-phosphate and NADH, the first reaction unique

20

25

180

232/116

to GMP biosynthesis. Disruption of PUR5 is lethal.

*Pyrimidine Biosynthesis is essential for cell proliferation*

Pharmacologic inhibitors of pyrimidine biosynthesis are cytotoxic. These include  
5 drugs like phosphonacetyl-L-aspartate (PALA) which inhibits aspartate  
transcarbamylase, a key enzyme in de novo pyrimidine synthesis. Also, there is  
evidence of the essentiality of pyrimidine biosynthesis from yeast. For example, the  
saccharomyces cerevisiae CDC8 gene encodes thymidylate kinase, required for  
synthesis of dTTP. Disruption of CDC8 is lethal.

10

*DNA synthesis and polymerization.*

Cell division clearly requires DNA polymerization to replicate the chromosomes so  
that each daughter cell has the same genetic makeup as the parent cell. Much of the  
15 basic machinery of DNA replication is conserved in prokaryotic and eukaryotic cells  
(1). Disruption of genes that encode proteins of DNA replication in yeast - including  
Polymerases I and III (the counterparts of human polymerases  $\alpha$  and  $\delta$ ), and accessory  
factors such as Replication Protein A and Replication Factor C - is lethal in *S.*  
*cerevisiae* (2). Nucleotide analogs that are incorporated into DNA are cytotoxic drugs.  
20 Examples of such analogs are the antineoplastic drug 6-mercaptopurine and arabinosyl  
NTPs, which interfere with DNA polymerization. Since inability to replicate DNA is  
lethal for growing cells, mutants in DNA replication must be obtained as conditional  
lethals in both prokaryotes and eukaryotes.

20

25

Second strand DNA polymerization on takes place in three main steps, each requiring  
different protein machinery: (1) At the start of replication an initiation complex is  
formed at chromosome structures called origins of replication. The parental DNA  
strands are transiently separated, a replication fork is formed and DNA synthesis is  
primed. (2) The elongation phase of replication is thought to take place in two

complexes, one moving forward on the leading strand and the other moving iteratively in the opposite direction to form the lagging strand. Elongation, then, requires replicative DNA polymerases and associated factors for unwinding and transiently stabilizing single stranded DNA, proofreading the newly synthesized template and, on the lagging strand, removing RNA primers and covalently linking adjacent newly synthesized lagging strands (Okazaki fragments). (3) During the final phase of DNA synthesis replication is terminated and the newly synthesized strands are separated.

Origin recognition complexes are formed by at least 6 origin recognition complex proteins (ORC 1 through 6) along with other factors, including "licensing" proteins such as the MCM family as well as "regulating" factors. The two principal nuclear replicative polymerases are DNA polymerase  $\alpha$ , which is responsible for priming synthesis and for synthesis of the lagging strand, and DNA Polymerase  $\delta$ , which synthesizes the leading strand. Both are multisubunit proteins, which function in multiprotein assemblies that include Replication Protein A, Replication Factor C, Proliferating Cell Nuclear Antigen and other proteins.

DNA Polymerases  $\beta$  and  $\epsilon$  are believed to principally carry out nuclear repair synthesis, while Polymerase  $\gamma$  is the mitochondrial replicative enzyme. These polymerases are also multiprotein complexes.

Proteins such as topoisomerases I and II and other DNA helicases are also required during replication to maintain DNA topology.

The biochemistry of replication termination is not well characterized however the proteins which carry out this final step of replication are likely to be essential.

#### *Inhibitors Have Been Identified for Several Steps of DNA Replication*

In addition to lethal disruptions of genes encoding proteins required for replication, a variety of cytotoxic inhibitors of DNA replication have been identified. They include



agents which act on production of DNA precursors as well as inhibitors of DNA polymerases.

*DNA Replication Inhibitors are Cytostatic or Cytotoxic Drugs*

5

There are several chemotherapy drugs that arrest DNA replication and poison cells by inhibiting production of deoxynucleotides, the precursors of DNA. These drugs include hydroxyurea, which inhibits ribonucleotide reductase, and 5-fluorouracil, which inhibits thymidylate synthase. Other inhibitors of replication appear to act, at least in part, by blocking DNA polymerases. These include nucleotide analogs that block DNA polymerases, such as 2',3' dideoxy NTPs and 3' deoxy ATP (cordycepin) as well as the chemotherapy drugs cytarabine (cytosine arabinoside), fludarabine phosphate and 2-chlorodeoxyadenosine. Cytarabine, after metabolism to the di- and trinucleotide phosphate forms, is incorporated into DNA and inhibits chain elongation leading to cell death, apparently by inducing apoptosis. Fludarabine, after metabolism to the triphosphate derivative, inhibits DNA polymerase, DNA primase and ribonucleotide reductase and is incorporated into DNA and RNA (3).

10

15

*DNA polymerization is essential for cell proliferation*

20

The essentiality of the function of DNA polymerization is clear, as such polymerization is needed for cell division, and therefore for tissue or tumor growth. As indicated for other categories, confirmation of the essentiality of a particular gene and the presence of a single locus, along with the determination of appropriate LOH and sequence variance heterozygosity characteristics identifies or confirms a gene in this category as an appropriate gene for potential allele specific targeting.

25

*References*

1. O'Donnell, M., Onrust, R., Dean, F.B., Chen, M. and J. Hurwitz (1993) Homology in accessory proteins of replicative polymerases - E. coli to humans. *Nucleic Acids Research* 21:1-3.
2. Stillman, B. (1996) Cell Cycle Control of DNA Replication. *Science* 274: 1659-1664.
3. Gandhi, V., Huang, P. and W. Plunkett (1994) Fludarabine inhibits DNA replication: a rationale for its use in the treatment of acute leukemias. *Leukemia and Lymphoma* 14 Suppl. 2: 3-9.

#### *Maintaining RNA at levels required for cell growth or survival*

Gene transcription is necessary for the production of messenger RNAs, the precursors of all cellular proteins. Transcription is also required for the production of ribosomal RNA, essential to formation of ribosomes, and for the production of transfer RNA, required for formation of aminoacyl tRNAs, the building blocks of protein synthesis. Turning off transcription - which can be accomplished with drugs that act on DNA templates or RNA polymerase - leads to rapid arrest of cell growth and subsequent cell death. Beyond gene transcription lie a series of essential RNA processing steps, including, but not necessarily limited to, mRNA splicing, capping, polyadenylation and export to the cytoplasm. Interference with any of these steps prevents the production of mature mRNA competent for translation, and therefore has the same cytotoxic effects as blocking transcription.

#### *Gene Transcription and RNA Processing Require Many Proteins*

Transcription of eukaryotic genes is carried out by three different RNA polymerases, each of which works with a different set of accessory factors. RNA Polymerase I is responsible for transcription of ribosomal RNAs, RNA Polymerase II transcribes protein coding genes and RNA Polymerase III transcribes transfer RNAs and other small RNAs. All three polymerases are multiprotein complexes. Several protein subunits are common components of all three polymerases, but each polymerase also

has unique subunits and accessory factors, not all of which have yet been identified or characterized. Some of the key proteins identified so far are:

RNA Polymerase I subunits and accessory factors including UBF1 and SL1. (SL1 has been shown to consist of TATA binding protein and three TATA associated factors.)

RNA Polymerase III subunits and accessory factors including TFIIIA, TFIIIB and TFIIIC.

RNA Polymerase II and its accessory factors are by far the most extensively characterized and most complex system. The large multisubunit protein complex that transcribes protein coding genes has recently come to be called the RNA Pol II holoenzyme (reviewed by Berk, ref. 1). The holoenzyme consists of more than 50 proteins, among which are:

RNA polymerase, the catalytic complex at the core of the holoenzyme. It consists of 14 subunits, many of which can complement their yeast counterparts *in vivo*.

The general transcription factors. These are proteins which either make direct contact with DNA, like TATA binding protein and associated factors, or interact with other transcription factors and/or transcriptional regulators. The general transcription factors, including TFII A, TFII B, TFII D, TFII E, TFII F, TFII H and TFII I, are multimeric protein complexes with >30 protein constituents (2,3). For example, there are 8-13 proteins which associate with TATA binding protein (called TATA associated factors, or TAFs) to collectively make up TFII D. Some of these factors (e.g. TFII250) have already been proven essential for cell proliferation.

Accessory proteins such as elongation factors, termination factors, activator and mediator proteins, srb (suppressor of RNA Polymerase B; see ref. 1 and references therein) proteins, RNA methylases and a variety of other processing factors.

RNA helicases, which are required for proper folding of RNAs,

Once transcribed, genes are spliced by multiprotein assemblies termed spliceosomes (4), which are made up of pre-mRNA, small nuclear ribonucleoproteins including (snRNPs) U1, U2, U4/6 and U5 and other proteins including SF2/ASF, U2AF and SC35. Recently progress has been made in cloning cDNAs for several splicing factors, however many of the proteins which process mRNAs have not yet been well characterized. After splicing, mRNAs are polyadenylated and exported to the cytoplasm (5). Several of the proteins of polyadenylation have been purified and cloned. The export of mRNAs is less well studied but is clearly a specific process requiring protein machinery. Several essential yeast genes required for mRNA transport have been identified.

*Inhibitors Have Been Identified for Several Steps of Gene Transcription*

The best studied inhibitors of gene transcription are small molecules that inhibit RNA polymerase or interact with DNA to block transcription. Inhibitors of RNA polymerase include actinomycin D, which intercalates into double stranded DNA and blocks the movement of RNA polymerase and rifampicin, an antibiotic which binds the  $\beta$  subunit of *E. Coli* RNA polymerase and blocks initiation of transcription. The best studied specific inhibitor of eukaryotic RNA Polymerase II is the potent mushroom toxin  $\alpha$ -amanitin, a cyclic octapeptide which binds to the polymerase with high affinity ( $K_d \sim 10^{-9}$  M). Several mutations conferring resistance to  $\alpha$ -amanitin have been characterized and they all map to the RNA Polymerase II protein coding sequence.

*Examples of essential yeast genes (disruption shown to be lethal) required to maintain cellular nucleic acids at levels compatible with cell growth or survival*

A number of yeast genes involved in DNA (including nuclear DNA and mt DNA) and RNA (including mRNA, tRNA and rRNA) metabolism have been disrupted and shown

essential for yeast cell viability. Many of these genes are conserved in all eukaryotes. Human homologs of these yeast genes are likely to be essential for human cell growth or survival. Specific examples:

5       The yeast DNA2 gene encodes a DNA helicase required for DNA replication. DNA2 is essential to the function of TOP2 (topoisomerase) which is also an essential gene.

10       POL1, POL2, POL3 and POL12 encode DNA polymerases. The disruption of any one of these genes is lethal. Knockout of polymerase associated genes DBP2 and POB3 is also lethal. These genes are essential for the synthesis of DNA.

15       ORC1, ORC2, ORC3, ORC4, ORC5, ORC6, CDC7, CDC46 and CDC54 are essential in yeast. These genes encode origin recognition complex proteins responsible for the initiation of DNA synthesis. There are direct human homologs of the ORC genes.

20       General replication factors RFA1, RFA2, RFA3, RFC1, RFC2, RFC3, RFC4 and RFC5 are all essential yeast genes. These genes encode replication protein A and replication factor C which are essential for DNA replication and have direct human homologs.

25       TBF1, TEL2 and CDC13 are essential yeast genes that encode proteins that responsible for the synthesis and maintenance of telomeres.

30       RNR1 (Ribonucleotide Reductase 1), RNR2 (Ribonucleotide Reductase 2) CDC8 (Thymidylate Kinase) and PUR5 (Inosine-5'-monophosphate dehydrogenase) are essential yeast genes involve in the purine/pyrimidine biosynthesis pathways and in the conversion of ribonucleotides to deoxyribonucleotides.

35       ROX3, RPA135, RPA190, RPA43, RPB10, RPB11, RPB2, RPB3, RPB5, RPB6,

RPB7, RPB8, RPC10, RPC128, RPC19, RPC25, RPC31, RPC34, RPC40, RPC53, and RPC82 are subunits of RNA polymerases I, II and III. These genes have been disrupted and shown to be essential. RNA polymerase I, II, and II are responsible for the synthesis of rRNA, mRNA, and tRNA respectively and have human homologs.

5

BRR2, DBP5, DBP6, DED1, HCA4, MAK5, and ROK1 are RNA helicases that are essential for processes such as pre-mRNA splicing and ribosomal RNA splicing.

10

Yeast TATA binding proteins TAF145, TAF17, TAF19, TAF25, TAF40, TAF47, TAF47, TAF60, TAF61, TAF67, and TAF90 are required for mRNA transcription by the RNA Polymerase II holoenzyme.

Transcription elongation factors RPO21 and RPO31 are essential.

15

General transcription factors SPT15, SSL1, SSL2, SUA7, TFA1, TFA2, TFB1, TFB2, TFB4, TFC2, TFC3, TFC4, TFC5, TFG1, TFG2, TOA1, and TOA2 have been disrupted and proven to be essential. These genes encode proteins that constitute the general machinery of RNA transcription.

20

Specific transcription factors BBP1, BRF1, BUR6, CDC39, HSF1, KIN28, MET30, RAP1, and REB1 are essential yeast genes. These genes encode proteins that are involved in the transcription of specific genes.

25

CUS1, GIN10, MSL5, PRP19, PRP31, SLU7, SME1, SNP2, USS1, and YHC1 are essential genes responsible for normal RNA splicing.

ESS1, FIP1, NAB2, NAB3, NAB4, PAP1, RNA14, RNA15, and YTH1 are essential genes required for RNA modification. The encoded proteins perform functions such as cleavage and polyadenylation of 3' ends of RNAs to produce mature mRNA

molecules.

### *References*

1. Berk, A.J. (1995) Biochemistry meets genetics in the holoenzyme. *Proc. Natl. Acad. Sci. U.S.A.* 92: 11952-11954.
2. Buratowski, S. (1994) The basics of basal transcription by RNA polymerase II. *Cell* 77: 1-3.
3. Tjian, R. and T. Maniatis (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77: 5-8.
4. Green, M.R. (1991) *Annual Review of Cell Biology* 7: 559.
5. Proudfoot, N. (1991) Poly(A) signals. *Cell* 64: 671-674.

### **Example 6. Genes required to maintain integrity and function of cellular and subcellular structures compatible with cell growth or survival.**

In order to survive and grow cells must be able to maintain their shape and internal architecture, including the structural integrity of a wide variety of subcellular organelles including the nucleus, mitochondria, endoplasmic reticulum and Golgi vesicles and a variety of lysosomes, peroxisomes vesicles and vacuoles. These structures perform essential functions such as:

- (i) Movement of proteins and other macromolecules across membranes to maintain their concentration at levels compatible with cell growth or survival. Newly synthesized proteins are transported to the endoplasmic reticulum by specialized transport apparatus which assists in protein folding and posttranslational modification. From the ER, proteins may be transported to distant cellular sites via vesicles which are comprised of specialized proteins. Some proteins synthesized in the cytoplasm must be transported into the mitochondria for proper mitochondrial function. There

also exist specialized apparatus for transport of mRNA from the nucleus.

(ii) Fusion or fission of various membrane bound cytoplasmic or nuclear organelles requires the specialized function of molecules that affect membrane properties to allow joining or separating and that provide a scaffold for moving membrane bound structures together or apart. The relationship of the ER and Golgi vesicles involves a continuous process of fission, while various classes of vacuoles or vesicles may fuse.

(iii) There must be effective coordination of the function of all cellular compartments. Coordination is accomplished by the transmission of signals from membrane to nucleus, from cytosol to mitochondria, from nucleus to cytoplasm, etc. Signals are transmitted by enzymes such as adenylate cyclases, protein kinases and protein phosphatases.

(iv) Maintenance of the integrity of cellular and subcellular structures also requires processes and structures for eliminating, transforming, sequestering or otherwise regulating levels of endogenous cellular toxins or waste substances. This may be accomplished by transfer of waste molecules to organelles such as vacuoles, lysosomes or peroxisomes, by inactivation of toxic byproducts of oxygen metabolism such as free radicals or by export of molecules that have reached excessive levels in the cell.

(v) The structure of the cytoplasm is maintained by the cytoskeleton, while different organelles in some cases are made up of specialized structural molecules. For example, the nucleus, bound by a double layered nuclear envelope, contains the nuclear matrix, consisting of over 100 unique proteins, as well as the histones and other proteins which form chromatin and the proteins which form subnuclear structures such as nucleoli, nuclear pores and the protein structures which convey mRNA out of the nucleus. (Darnell, J. et al., Molecular Cell Biology, Scientific American Books, 1990.)

The fibrous proteins of the cytoplasm are collectively referred to as the cytoskeleton. Among the important cytoskeletal proteins are microfilaments made up of actin molecules, microtubules made up of tubulin molecules, and intermediate filaments, made up of one of a variety of subunit types. The cytoskeleton is important not only



for maintenance of cell shape, strength and rigidity but also for providing a frame for movement of other structures. Microtubules, for example, are critical for chromosome movement during cell division, while actin microfilaments and intermediate filaments affect the organization and mobility of surface membrane proteins. Actins and other cytoskeletal proteins are vital for processes such as endocytosis, which is the only route of essential nutrients such as transferrin-bound iron. Cells also contain a variety of proteins essential for anchoring organelles to the cytoskeleton, or anchoring the plasma membrane to adjacent structures such as basement membranes and adjacent cells.

5

10

A variety of yeast structural proteins required to maintain integrity and function of cellular and subcellular structures have been disrupted and shown essential for cell survival. Since most structural proteins are highly conserved in eukaryotes it is likely that the human counterparts of these yeast genes are also essential. Specific examples:

15

The genes encoding yeast nuclear pore proteins (nucleoporin) NIC96, NSP1, NUP49, NUP57, NUP82, NUP145 and NUP159 are lethal when disrupted, as is the pore trafficking protein GSP1. NNF1 is an essential protein of the nuclear envelope required for proper nuclear morphology.

20

The yeast nucleolar protein NOP2, homologous to human proliferation associated nucleolar antigen p120, is essential. NOP4 encodes another essential yeast nucleolar protein.

25

Knockout of the yeast ACT1 gene, which encodes actin, is lethal, as is knockout of the actin related proteins ARP100, ARP2, ARP3 and ARP4. The actin binding and severing protein cofilin, encoded by the yeast COF1 gene, is also essential, as is profilin (PFY1), which can complex with actin monomers and prevent actin polymerization. PAN1 is an essential protein involved in normal regulation of the actin cytoskeleton.

191

232/116

The RET1, RET2, RET3, SEC1, SEC4, SEC5, SEC6, SEC7, SEC8, SEC10, SEC11, SEC14, SEC16, SEC17, SEC18, SEC20, SEC26, SEC27, SEC31, SEC61, SEC62, SEC63, SFT1, SLY1, BET1, BET3, UFE1, USO1, VTI1, TIP20, KAR2 and BOS1 genes are all essential in yeast. These genes encode proteins which are vital for the function of the endoplasmic reticulum and Golgi vesicles, including processes such as protein transport across the ER, membrane fusion and formation of vesicles.

The essential yeast histone-like protein CSE4 is required for chromosome segregation. STH1, RSC6 and RSC8 are components of the essential abundant chromatin remodeling complex, while SPT5 and SPT6 influence gene expression through effects on chromatin structure.

The essential yeast intermediate filament protein MDM1 is involved in organelle inheritance and mitochondrial morphology.

The essential yeast mitochondrial proteins MGE1 and SSC1 participate in folding of proteins during mitochondrial import. TIM17, TIM22 and TIM23 are essential mitochondrial inner membrane proteins involved in import and translocation of proteins. ATM1 is an ATP binding mitochondrial inner membrane protein.

The RAT1, MTR2 and MTR3 genes encode proteins essential for mRNA transport from the nucleus to the cytoplasm.

DBF8 is an essential yeast protein involved in protein transport.

APS2 is an essential clathrin associated protein, involved in membrane transport.

The yeast PKC1 gene encodes the essential protein kinase C, which regulates the MAP kinase cascade; CDC15 is an essential component of the MAP kinase kinase kinase

192

232/116

family of signaling proteins.

CYR1 is an essential adenylate cyclase which generates cAMP in response to signals including ras activation. GDI1 is an essential GDP dissociation inhibitor.

5

#### **Example 8: Validation of Target Gene Essentiality**

To investigate whether specific target genes are essential for cell proliferation and/or survival, a method was developed to use antisense oligonucleotides to inhibit gene expression. Phosphorothioate antisense oligonucleotides targeting polymorphic sites were transfected into human cell lines, and mRNA down-regulation was assessed by northern blotting. mRNA down-regulation was achieved for 19 of the 35 polymorphisms targeted (54.2%). Oligonucleotides targeting each polymorphic allele were (separately) transfected to assess the allele-specificity of the mRNA down-regulation. In 15 of the 19 sites accessible to oligos, the oligonucleotide targeting the allele found in the cell down-regulated mRNA to a level significantly lower than did the mismatched oligonucleotide. In 6 of these 15 cases, striking allele-specificity was observed.

The consequence of down-regulating the mRNA of an essential gene should be cell death. Allele-specific cell death was indeed observed in these experiments, both upon transfecting cells daily for three days with the phosphorothioate oligos described above (followed by a recovery period during which control-treated cells continued to divide while essential gene inhibition prevented division) or upon extended (5-10 days) daily transfections with less toxic oligonucleotide chemistries. In an experiment targeting either RNA polymerase II or the glutamyl-prolyl tRNA synthetase (EPRS), cells were transfected for five consecutive days with oligos either targeting the allele found in the cell (match) or carrying a 1 bp mismatch, targeting the other allele (mismatch). One day after the fifth transfection, cells remaining on the plate were quantitated by staining

with sulforhodamine B. The matched oligonucleotide was significantly more cytotoxic than the mismatched oligonucleotide.

5       **Example 9: Aminoacyl tRNA Synthetases are Essential for Cell Survival**

Each aminoacyl-tRNA synthetase performs an analogous role in protein synthesis, and each represents a target for the present invention.

10       *Aminoacyl-tRNA synthetases perform a basic cell function*

Aminoacyl-tRNA synthetases are present in all living cells (1). (A recent paper entitled "A minimal gene set for cellular life derived by comparison of complete bacterial genomes" [ref. 2] concludes that as few as 256 genes may be required for prokaryotic cell life; all 20 tRNA synthetases are included in this minimal gene set.) Each tRNA synthetase catalyzes ATP dependent covalent attachment of a specific amino acid to its cognate tRNA. It is the specificity of each synthetase for a single amino acid and transfer RNA that establishes the universal rules of the genetic code. The aminoacyl-tRNAs produced by tRNA synthetases constitute the precursors for protein assembly by ribosomes - thus tRNA synthetases are vital for peptide polymerization and processing. Blockade of peptide polymerization and processing at any one of multiple different steps (see above) results in arrest of cell growth and eventually cell death in a variety of organisms and cell types.

25       *Aminoacyl-tRNA synthetases have been shown essential in all tested organisms*

It has been demonstrated by mutagenesis experiments that tRNA synthetases are essential for prokaryotic, yeast and mammalian cell survival (ref. 1-5). The most relevant data concerns mammalian cells: mutagenesis of Chinese hamster ovary (CHO)

and Chinese hamster lung cells followed by "suicide" selection at 39°C for temperature sensitive (ts), conditionally lethal protein synthesis mutants has led to isolation of cell lines with mutant tRNA synthetases (reviewed in ref. 5). (The "suicide" of dividing cells is accomplished by adding thialysine or tritiated [3H] amino acids to cell growth media. Only cells that incorporate these amino acid analogs into protein die - thus cells that are protein synthesis deficient at 39°C survive the selection.) The fraction of cells surviving a single round of suicide selection ranges from one in 105 to one in 108. Biochemical and genetic characterization of surviving cells has led to identification of specific ts aminoacyl-tRNA synthetase mutants. Cell lines with mutant leucyl- or asparaginyl-tRNA synthetases have been isolated repeatedly because the genes for leu and asn tRNA synthetases are haploid in the CHO cell line used for selection, and therefore require only one mutation. Less frequently, mutant alanyl-, arginyl-, glutaminyl-, histidyl-, lysyl-, methionyl-, tryptophanyl- and valyl-tRNA synthetases have been isolated. The properties of these mutant cell lines are similar: when shifted to 39°C, the non-permissive temperature, the rate of protein synthesis drops, in some cases to almost undetectable levels. Soon thereafter the cells stop replicating DNA and within a few days cell death ensues. These experiments constitute proof of the essential role of tRNA synthetases in mammalian cells. Arrest of protein synthesis and consequent cell death can be prevented in some cases by supplementing cell media with the amino acid substrate of the defective tRNA synthetase (thereby driving the aminoacylation reaction), or by fusing the mutant cell line with a normal cell line, or a cell line mutant for a different tRNA synthetase (thereby complementing the mutant synthetase). The cell fusion experiments show that the aminoacyl-tRNA synthetase mutations are recessive at the cellular level. The chromosomal map positions of a number of human tRNA synthetases were first determined by analysis of (human) X (ts mutant CHO cell) hybrids. Human chromosomes are progressively lost in such hybrids, but one human chromosome - the one which contains the human synthetase complementary to the mutant hamster synthetase - is consistently retained. Such experiments provided the first evidence that

human tRNA synthetases are single copy genes in man (or at least confined to a single chromosome; refs. 6, 7). Subsequently Southern blotting and fluorescence *in situ* hybridization analyses have confirmed and extended these observations for thirteen synthetases (8-14). These Southern blotting and *in situ* hybridization mapping studies established beyond doubt that each of the human tRNA synthetase genes investigated is encoded at a single locus. The table below summarizes the chromosomal location of tRNA synthetases mapped to date.

#### Chromosome Location of tRNA Synthetases

tRNA synthase	Chromosome	tRNA Synthetase	Chromosome
Ala	16q22	Trp	14q21-32
Arg, Leu, His, Thr	5	Asp	2
Asn	18	Gln	3p
Cys	11p15.5	Gly	7
Glu/Pro	1q32-42	Ile (mitochondrial)	2
Gly	7p15	Lys	16q21
Ile	9q21	Ser	1p12
Lys	16q23-24	Tyr	1p31
Met	12	Val	6p21.3 9

#### *Classification of tRNA synthetases*

The twenty tRNA synthetases are divided into two groups based on structural features and functional properties that are conserved throughout evolution. There are ten class I synthetases, all of which contain two short conserved amino acid segments which fold together to form an ATP binding pocket called the Rossman fold, in the amino

terminal half of the proteins. The C-terminal end of the Class I synthetases contains the tRNA binding fold. Class II synthetases, of which there are also ten, share up to three short conserved amino acid motifs.

5

## References

1. Schimmel, P., Giege, R., Moras, D. and S. Yokoyama (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.* 90:8763-8768.
- 10 2. Mushegian, A.R. and E.V. Koonin (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U.S.A.* 93:10268-10273.
3. Schmitt, E. and P. Schimmel (1994) Dominant lethality by expression of a catalytically inactive class I tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.* 90:6919-15 6923.
4. Ripmaster, T.L., Shiba, K. and P. Schimmel (1995) Wide cross-species aminoacyl-tRNA synthetase replacement in vivo: Yeast cytoplasmic alanine enzyme replaced by human polymyositis serum antigen. *Proc. Natl. Acad. Sci. U.S.A.* 92:4932-4936.
5. Wasmuth, J.J., Chinese Hamster Cell Protein Synthesis Mutants, pp. 375-421 in 20 Gottesman, M., editor, *Molecular Cell Genetics*, John Wiley & Sons, New York, 1985.
6. Kunze, N., Bittler, E., Fett, R., Schray, B., Hameister, H., Wiedorn, K.-H. and R. Knippers (1990) The human QARS locus: assignment of the human gene for glutaminyl-tRNA synthetase to chromosome 1q32-42. *Human Genetics* 85:527-530. (This gene was later shown to be glutamyl-prolyl tRNA synthetase [abbreviated 25 EPRS]).
12. Cirullo, R.E., Arredondo-Vega, F.X., Smith, M. and J.J. Wasmuth (1983) Isolation and characterization of interspecific heat-resistant hybrids between a temperature-sensitive Chinese hamster cell asparaginyl-tRNA synthetase mutant and normal human leukocytes: assignment of human asnS gene to chromosome 18. *Somatic Cell Genetics*

9:215-233.

13. Nichols, R.C., Pai, S.I., Ge, Q., Targoff, I.N., Plotz, P.H. and P. Liu (1995) Localization of two human autoantigen genes by PCR screening and *in situ* hybridization - glycyl tRNA synthetase locates to 7p15 and alanyl-tRNA synthetase locates to 16q22. *Genomics* 30:131-132.

14. Nichols, R.C., Blinder, J., Pai, S.I. et al. (1996) Assignment of two human autoantigen genes: isoleucyl tRNA synthetase locates to 9q21 and lysyl-tRNA synthetase locates to 16q23-24. *Genomics*: 210-213.

#### **Example 11: Sodium Potassium ATPase, 1 subunit (ATP1A1) - Target Gene VARIA125**

##### *Sodium Potassium ATPase is essential for cell survival*

The plasma membranes of virtually all eukaryotic cells contain a Na<sup>+</sup>, K<sup>+</sup> pump that operates as an antiport, pumping Na<sup>+</sup> out of the cell and K<sup>+</sup> in against their concentration gradients. In coupling the hydrolysis of ATP to the active transport of 3 Na<sup>+</sup> out and 2 K<sup>+</sup> into the cell the pump is electrogenic. The electrochemical gradients generated and maintained by the Na<sup>+</sup>,K<sup>+</sup> pump are essential for the regulation of cell volume, and for the secondary, sodium-coupled active transport of a variety of organic and inorganic molecules including glucose, amino acids and Ca<sup>++</sup>. Hence the sodium potassium pump plays an essential role in cellular physiology (1).

Sodium Potassium ATPase is a heterodimer composed of a ~100 kDa catalytic subunit and a ~55 kDa glycoprotein subunit of unknown function. Biochemical studies and gene cloning have demonstrated the existence of three isoforms and two -like isoforms of the catalytic subunit, each encoded by a separate gene and with a characteristic expression pattern (reviewed in refs. 2 and 3). Of these, only the 1 gene



(ATP1A1) is ubiquitously expressed; the other subunits have restricted tissue distribution.

Sodium Potassium ATPase is the target of the cardiac glycoside drugs, including digoxin and the poison ouabain. Ouabain binds to the extracellular face of the  $\alpha$  subunit and inhibits  $\text{Na}^+$ ,  $\text{K}^+$  exchange, leading to cell death. The  $\alpha$  subunit from primates is sensitive to nanomolar concentrations of ouabain while the rodent  $\alpha$  subunit is resistant to ~1000 fold higher concentrations, enabling precise definition of the ouabain binding site. Study of human-rat chimeric  $\alpha$  subunits combined with site directed mutagenesis has localized the ouabain interacting domain in the amino-terminal portion of the  $\alpha$  subunit (4,5). Other structure-function studies have contributed to an understanding of  $\alpha$  subunit cation binding and ATPase functions, while electron microscopy and low resolution (20-30 Å) diffraction analyses of membrane preparations have elucidated the geometry of the protein in the membrane (1).

*The  $\alpha$  subunit of Sodium Potassium ATPase has sequence variants*

The cDNA sequence of the human  $\alpha$  subunit of sodium-potassium ATPase has been published by four groups (6-9). We undertook a systematic search for DNA sequence variance by analyzing the  $\alpha$  cDNA from 36 unrelated individuals using the single strand conformation polymorphism technique. Primers were designed using the sequence of Kawakami et al. (GENBANK accession D00099; see ref. 6). SSCP analysis revealed 7 sequence variances, and subsequent DNA sequence analysis confirmed that nucleotides 1059 (A vs. C), 1428 (G vs. A), 2538 (T vs. C), 3324 (C vs. T), 3375 (G vs. A), 3397 (G vs. A) and 3408 (C vs. A) vary as shown in the Target Summary Table. The first five sequence variances are in the coding sequence while the latter two are in the 3' untranslated region.

The frequency of heterozygotes for the seven sequence variants ranged from 3-11% among the 36 individuals tested. Some of the sequence variances appear to occur more commonly in certain racial or ethnic groups. For example, heterozygotes for four sequence variances (at nucleotides 1059, 1428, 3324 and 3375) were detected solely or predominantly in North American Blacks, with heterozygote frequencies of 1/4 or 2/4. The nucleotide 2538 variance was detected solely in North American Whites (4/16) and results in an amino acid exchange (see below). The nucleotide 3397 sequence variance was detected solely in one Japanese individual (of four tested).

The nucleotide 2538 sequence variant results in an aspartic acid vs. glutamic acid substitution at amino acid 740 of the 1024 amino acid protein. This residue lies in the cytoplasmic loop of the 1 subunit.

*The alpha1 subunit of Sodium Potassium ATPase maps to chromosome 1p13-p11*

The gene for the 1 subunit of sodium-potassium ATPase has been mapped to chromosome band 1p13-p11 by several techniques. Yang-Feng et al. (10) assigned the ATP1A1 gene to 1p21-cen by Southern analysis of DNA from panels of rodent/human somatic cell hybrid lines. This localization was confirmed and refined by Chehab et al., who showed that the gene for the ATP1A1 subunit is on 1p13-p11 using hybridization to flow-sorted chromosomes and *in situ* hybridization (9).

*Chromosome band 1p13-p11 is a site of frequent loss of heterozygosity*

The short arm of chromosome 1 is comparatively well investigated for allele loss, especially in breast and colon cancers, however most of these studies are principally concerned with the 1p36 region, and there is comparatively little data on 1p13-p11. The best studies of proximal 1p allele loss are in breast and testicular cancers. These studies show LOH occurs in approximately 15-35% of breast cancers (11,12) and 15-25% of testicular cancers (13). Data from more distal loci on 1p show >25% LOH in

glioma, colon cancer, stomach cancer, ovarian cancer, and liver cancer (14). The LOH observed in this region indicates that other essential genes mapping to the 1p chromosomal arm, and especially to the 1p11 region, which have LOH and for which sequence variances, and therefore heterozygotes for a sequence variance, exist in normal somatic cells of individuals in a population are potential target genes

### References

1. Jorgensen, P.L. Na, K-ATPase, structure and transport mechanism. In De Pont, *ed. Molecular Aspects of Transport Proteins*, Elsevier Science Publishers, The Netherlands, 1992, pp. 1-26..
2. Sweadner, K.J. (1989) *Biochimica et Biophysica Acta* 1154: 185-220.
3. Lingrel, J.B., Orlowski, J., Shull, M.M. and E.M. Price (1989) *Prog. Nucleic Acid Research and Mol. Biol.* 38: 37-89.
4. Price, E.M. and J.B. Lingrel (1988) Structure-function relationships in the Na, K-ATPase a subunit: site directed mutagenesis of glutamine-111 to arginine and asparagine 122 to aspartic acid generates a ouabain-resistant enzyme. *Biochemistry* 27: 8400-8408.
5. Emanuel, J.R., Graw, S., Housman, D. and R. Levenson (1989) Identification of a region within the Na, K-ATPase a subunit that contributes to differential ouabain sensitivity. *Molecular and Cellular Biology* 9: 3744-3749.
6. Kawakami, K., Ohta, T., Nojima, H., Nagano, K. (1986) Primary structure of the alpha-subunit of human Na,K-ATPase deduced from cDNA sequence. *J. Biochem.* 100: 389-397.
7. Ovchinnikov, Y. A., Monastyrskaya, G. S., Broude, N. E., et al. (1987) The family of human Na<sup>+</sup>,K<sup>+</sup>-ATPase genes: a partial nucleotide sequence related to the alpha-subunit. *FEBS Lett.* 213: 73-80.
8. Shull, M. M. and J.B. Lingrel (1987) Multiple genes encode the human Na<sup>+</sup>,K<sup>+</sup>-ATPase catalytic subunit. *Proc. Nat. Acad. Sci. U.S.A.* 84: 4039-4043.

9. Chehab, F. F., Kan, Y. W., Law, M. L., Hartz, J., Kao, F.-T. and R. Blostein (1987) Human placental Na<sup>+</sup>,K<sup>+</sup>-ATPase alpha subunit: cDNA cloning, tissue expression, DNA polymorphism, and chromosomal localization. *Proc. Nat. Acad. Sci. U.S.A.* 84: 7901-7905.

5 10. Yang-Feng, T.L., Schneider, J.W., Lindgren, V., Shull, M.M., Benz, E.J., Jr., Lingrel, J.B. and U. Franke (1988) Chromosomal localization of human Na<sup>+</sup>,K<sup>+</sup>-ATPase alpha- and beta-subunit genes. *Genomics* 2: 128-138.

11. Bieche, I., Champeme, M.H., Matifas, F., Cropp, C.S., Callahan, R. and R. Lidereau (1993) Two distinct regions involved in 1p deletion in human primary breast cancer. *Cancer Res.* 53:1990-4.

10 12. Nagai H, Negrini M, Carter SL, et al. (1995) Detection and cloning of a common region of loss of heterozygosity at chromosome 1p in breast cancer. *Cancer Res.* 55:1752-7.

13. Mathew S., Murty V.V., Bosl G.J., Chaganti R.S.K. (1994) Loss of heterozygosity identifies multiple sites of allelic deletions on chromosome 1 in human male germ cell tumors. *Cancer Res.* 54:6265-9.

15 14. Yeh S.H., Chen P.J., Chen H.L., Lai M.Y., Wang C.C. and D.S. Chen (1994) Frequent genetic alterations at the distal region of chromosome 1p in human hepatocellular carcinomas. *Cancer Res.* 54:4188-92.

20  
**Example 12: Ribonucleotide Reductase, M1 subunit (RRM1) - Target Gene VARIA200**

25 *Ribonucleotide Reductase is essential for cell growth*

Human ribonucleotide reductase (also called ribonucleoside diphosphate reductase) is essential in dividing cells for the production of deoxyribonucleotides prior to DNA synthesis in S phase. Ribonucleotide reductase catalyzes the reduction of all four

ribonucleoside diphosphates to the corresponding deoxyribonucleoside diphosphates by replacing the 2' hydroxyl moiety of ribose with a hydride ion to form deoxyribose; these reactions constitute the first committed steps in the creation of DNA precursors (deoxyribonucleotides), and are therefore tightly regulated by allosteric nucleotide binding sites on the M1 subunit (2,3). The enzyme is an  $\alpha_2\beta_2$  tetramer apparently conserved in all prokaryotes and eukaryotes (1). The two subunits, M1 and M2, are both required for enzyme activity. The RRM2 subunit contains the catalytic site, while the RRM1 subunit provides an indispensable allosteric function. (See pages 758-763 of Biochemistry by C.K. Mathews and K.E. van Holde, Benjamin/Cummings Publishing Biochemistry, Company, Redwood City, 1990 for a fuller account of ribonucleotide reductase function.)

Both ribonucleotide reductase subunits are expressed in all proliferating cells but are generally nondetectable in quiescent cells. Ribonucleotide reductase subunit M2 is the target of several antineoplastic compounds, including hydroxyurea. Hydroxyurea is used in the chemotherapy of a variety of myeloproliferative disorders (4). It acts by reversibly destroying a tyrosyl free radical in the catalytic site of the M2 subunit (3). Hydroxyurea and other ribonucleotide reductase poisons are specific for the S phase of the cell cycle, resulting in growth arrest at the G1-S boundary and apoptotic death in tumor cells (5). Exposure of cell cultures to hydroxyurea results in selection of cells expressing high levels ribonucleotide reductase, demonstrating that ribonucleotide reductase is required for these cells to grow (6).

*The human ribonucleotide reductase gene has sequence variances*

The cDNA sequence of the human ribonucleotide reductase M1 subunit has been published by two groups (7,8). We undertook a systematic search for DNA sequence variance in the cDNA of the M1 subunit by analysing 36 unrelated individuals using the single strand conformation polymorphism technique. Primers were designed using

203

232/116

the sequence of Parker et al. (GENBANK accession X59543; see ref. 7). SSCP analysis revealed 4 sequence variances, and subsequent DNA sequence analysis confirmed that nucleotides 1037 (C vs. A), 2410 (A vs. G), 2419 (A vs. G) and 2717 (T vs. A) vary as shown in the Target Summary Table. (The sequence variance at nt 1037 was previously noted by Parker et al., ref. 7.) Also, DNA sequencing revealed an insertion/deletion sequence variance: the 9 consecutive T nucleotides between positions 2724 and 2732 (numbering from ref. 7) were augmented in some cDNAs by a tenth T. (This sequence variance is designated T9 vs. T10 in the Target Summary Table.)

Both alleles at nt 1037 were detected in North American Whites, Hispanics, Chinese, Japanese, Arabs and Indians. Similarly, both alleles of the sequence variance at nt 2410 were detected in virtually all tested populations: North American White, North American Black, Hispanic, Chinese, Arab and Indian. In contrast, the sequence variances at nt 2419 and 2717 were prevalent in North American Blacks, Hispanics, Chinese, and Japanese, but not North American Whites. The insertion/deletion sequence variance at nt 2724 was only studied in four individuals so no firm conclusions can be drawn regarding population distribution, however it appears to be in linkage disequilibrium with the 2419 and 2724 sequence variances.

*The human ribonucleotide reductase gene maps to chromosome 11p15.5*

The gene for human ribonucleotide reductase has been mapped to band 11p15.5 by several techniques. Initially the gene was localized by Southern hybridization analysis of human X rodent somatic cell hybrids and by chromosomal *in situ* hybridization (9). Subsequently RRM1 has been placed on a yeast artificial chromosome (YAC) physical map of chromosome 11p15 (10). The precise physical localization of the RRM1 gene facilitates interpretation of LOH results at adjacent polymorphic markers (see below).

*Chromosome band 11p15.5 is a site of frequent loss of heterozygosity*

The short arm of chromosome 11 is the site of several tumor suppressor genes, including the WT1 gene and the Beckwith-Weidemann syndrome gene. As a result there are many studies of LOH in 11p15.5, particularly focusing on breast, cervix, kidney, liver, lung, ovarian, stomach and testicular cancers. These studies show that the 11p15.5 band of chromosome 11 is frequently reduced to one copy (11-28). For example, LOH occurs in approximately 13-33% of breast cancers (11-13), 14-42% of cervical cancers (14), 0-50% of liver cancers (16), 0-80% of lung cancers (17-19), 18-54% of ovarian cancers (20,21), 0-71% of stomach cancers (22) and 0-50% of testicular cancers (23,24). Other studies show that 11p15.5 LOH may also be frequent in bladder cancer (25), esophageal cancer (26), some leukemias (27) and sarcomas (28). Many deletions in the 11p15.5 region span relatively short chromosomal segments (2 - 10 megabases; see ref. 17).

**References**

1. Caras, I.W., Levinson, B.B., Fabry, M., et al. (1985) Cloned mouse ribonucleotide reductase subunit M1 cDNA reveals amino acid sequence homology with Escherichia Coli and herpesvirus ribonucleotide reductases. *J. Biol. Chem.* 260: 7015-7022.
2. Thelander, L., and P. Reichard, (1979) Reduction of Ribonucleotides. *Annu. Rev. Biochem.* 48:133-158.
3. Reichard, P. and A Ehrenberg (1983) Ribonucleotide reductase: a radical enzyme. *Science* 221: 514-9.
4. Donehower, R.C. (1992) An Overview of the clinical experience with hydroxyurea. *Seminars in Oncology* 19:11-19, 1992.
5. Wright, P.S., Cross-Doersen, D., Thõng, J.P., et al. (1996) A ribonucleotide reductase inhibitor, MDL 101,731, induces apoptosis and elevates TRPM-2 mRNA levels in human prostate tumor xenografts. *Experimental Cell Research* 22: 54-60.

6. Cocking, J.M., Tonin, P.N., Stokoe, et al. (1987) Gene for M1 subunit of ribonucleotide reductase is amplified in hydroxyurea-resistant hamster cells. *Somat. Cell. Mol. Genet.* 13:221-33.
7. Parker, N.J., Begley, C.G. and R.M. Fox. (1991) The Human M1 Subunit of Ribonucleotide Reductase: cDNA Sequence and Expression in Stimulated  
5 Lymphocytes. *Nucleic Acids research* 9:3741.
8. Pavloff, N., Rivard, D., Masson, S., Shen, S.H. and A.M. Mes-Masson. (1992) Sequence Analysis of the Large and Small Subunits of Human Ribonucleotide Reductase. *DNA Sequence* 2:227.
9. Brissenden, J.E., Caras, I., Thelander, L. and Francke, U. (1988) The structural gene for the M1 subunit of ribonucleotide reductase maps to chromosome 11, band p15, in human and to chromosome 7 in mouse. *Exp. Cell. Res.* 174:302-8.
10. See: <http://shows.med.buffalo.edu/home.html>
11. Ali, I., Lidereau, R., Theilley, C. and R. Callahan (1987) Reduction to  
15 homozygosity of genes on chromosome 11 in human breast neoplasia. *Science* 238:185-8.
12. Winqvist, R., Mannermaa, A., Alavaikko, et al. (1993) Refinement of regional loss of heterozygosity for chromosome 11p15.5 in human breast tumors. *Cancer Research* 53: 4486-4488.
13. Carter, S.L., Negrini, M., Baffa, R., et al. (1994) Loss of heterozygosity at 11q22-  
20 q23 in breast cancer. *Cancer Research* 54:6270-4.
14. Mitra, A.B., Murty, V.V.V.S., Li, R.G., et al. (1994) Allelotype analysis of cervical carcinoma. *Cancer Research* 54:4481.
15. Fujimori, M., Tokino, T., Hino, O., et al. (1991) Allelotype study of primary  
25 hepatocellular carcinoma. *Cancer Research* 51: 89-93.
16. Wang, H.P. & C.E. Rogler (1988) Deletions in human chromosomes 11p and 13q in primary hepatocellular carcinomas. *Cytogenetics and Cell Genetics* 48:72-78.
17. Bepler, G. and Garcia-Blanco, M.A. (1994) Three Tumor Suppressor Regions on Chromosome 11p Identified by High Resolution Deletion Mapping in Human Non-



Small Cell Lung Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 91:5513-7.

18. Iizuka, M., Sugiyama, Y., Shiraishi, M., et al. (1995) Allelic losses in human chromosome 11 in lung cancers. *Genes, Chromosomes & Cancer* 13:40-46.

5 19. Weston, A., Willey, J.C., Modali, R., et al. (1989) Differential DNA sequence deletions from chromosomes 3, 11, 13 and 17 in squamous cell carcinoma, large-cell carcinoma and adenocarcinoma of the human lung. *Proc. Natl. Acad. Sci. U.S.A.* 86:5099-5103.

10 20. Kiechle-Schwartz, M., Bauknecht, T., Wienker, T., et al. (1993) Loss of Constitutional Heterozygosity on Chromosome 11p in Human Ovarian Cancer. *Cancer* 72:2423-32.

21. Viel, A., Giannini, F., Tumiotti, L., Sopracordevole, F., Visentin, M.C. and M. Boiocchi (1992) Chromosomal localization of two putative 11p oncosuppressor genes involved in human ovarian tumors *British Journal of Cancer* 66: 1030-1036.

15 22. Baffa, R., Negrini, M., Mandes, B., et al. (1996) Loss of heterozygosity for chromosome 11 in adenocarcinoma of the stomach. *Cancer Research* 56: 268-72.

23. Lothe, R.A., Hastie, N., Heimdal, K., et al. (1993) Frequent loss of 1p13 and 11p15 loci in male germ cell tumors. *Genes, Chromosomes & Cancer* 7:96-101.

24 Smith, R.C., and Rukstalis, D.B. (1995) Frequent Loss of Heterozygosity at 11p Loci in Testicular Cancer. *The Journal of Urology* 153:1684-7.

20 25. Shaw, M.E. and Knowles, M.A. (1995) Deletion Mapping of Chromosome 11 in Carcinoma of the Bladder. *Genes, Chromosomes & Cancer* 13:1-8.

26. Shibagaki, I., Shimada, Y., Wagata, T., et al. (1994) Allelotype analysis of esophageal squamous cell carcinoma. *Cancer Research* 54: 2996-3000.

25 27. Ahuja, H.G., Foti, A., Zhou, D.J. and M.J. Cline (1990) Analysis of proto-oncogenes in acute myeloid leukemia: loss of heterozygosity for the Ha-ras gene. *Blood* 75: 819-822.

28. Yamaguchi, T., Toguchida, J., Yamamuro, T., et al. (1992) Allelotype analysis in osteosarcoma: frequent allele loss on 3q, 13q, 17p and 18q. *Cancer Res.* 52: 2419.

**Example 13: Thymidylate Synthase (TS) - Target Gene VARIA250**

*Thymidylate Synthase is essential for cell growth*

5 Human thymidylate synthase (TS) catalyzes the formation of thymidine monophosphate (dTMP) from deoxyuridine monophosphate (dUMP) by transfer of a methyl group from N5,N10-methylenetetrahydrofolate to carbon 5 of dUMP. This is the sole *de novo* pathway to dTMP, an essential precursor for DNA synthesis. TS also plays an important role in balancing the four nucleotide precursors for DNA polymer synthesis (1). Thus TS is an attractive target for antiproliferative drugs. (See Biochemistry by C.K. Mathews and K.E. van Holde, Benjamin/Cummings Publishing Company, Redwood City, 1990, pages 763-768, for a fuller account of thymidylate synthase function.)

15 Like some other growth associated genes involved in DNA synthesis, thymidylate synthase is expressed in proliferating cells at 20-40 fold higher levels than in quiescent cells. Increased expression occurs at the G1-S transition of the cell cycle when quiescent cells are stimulated with serum. Levels of thymidylate synthase are finely controlled by autoregulatory feedback loops wherein TS protein regulates the transcription, stability and translational efficiency of TS mRNA (2). Transcription increases by only 2-4 fold, so posttranscriptional events constitute the predominant regulatory mechanisms (3). One mechanism of 5-FU resistance is increased expression of TS Mrna.

25 Thymidylate synthase is the target of 5-fluorouracil (5-FU), a potent antineoplastic compound. Once inside cells 5-FU is ribosylated and phosphorylated to 5-fluoro-2'-deoxyuridine 5'-monophosphate (F-dUMP), which acts as an inhibitory transition state analog of TS when bound in the presence of the enzyme's second substrate, N5,N10-methylenetetrahydrofolate. (5-FU is also incorporated into both DNA and RNA,

augmenting its toxicity.) 5-FU induces partial responses in 10-30% of patients with a variety of cancers, including metastatic breast and gastrointestinal tract cancers (4). While 5-FU is a potent antiproliferative agent in tissue culture cells, as with most antineoplastic drugs, its clinical utility is limited by lack of discrimination between  
5 normal cells and tumor cells: common toxic effects include stomatitis, diarrhea, bone marrow suppression, hair loss and occasionally cardiac and neurologic symptoms.

*The human thymidylate synthase gene has sequence variances*

10 The sequence of a human thymidylate synthase cDNA was determined by Takeishi et al. (5), who later determined the genomic sequence as well (6). We undertook a systematic search for DNA sequence variance by analysing 36 unrelated individuals using the single strand conformation polymorphism. Primers were designed using the sequence of Takeishi et al. (5). SSCP analysis revealed 3 DNA fragments having  
15 sequence variances, and subsequent DNA sequence analysis showed that nucleotides 1066 (C vs. T), 1136 (A vs. G) and 1497 (A vs. T) vary among normal individuals as shown in the Target Summary Table. All three sequence variances are in the 3' untranslated region of the gene. The nucleotide 1066 and 1497 sequence variances are in complete linkage disequilibrium in the 36 individuals examined. Both alleles of all  
20 three sequence variances were detected in North American Whites, North American Blacks, Chinese, Japanese, Arabs and Indians.

Another TS sequence variance has been described by Berger and colleagues (7-9). They detected a T to C change at nucleotide 276 of the TS gene, resulting in the  
25 substitution of histidine for an evolutionarily conserved tyrosine at residue 33 of TS protein. So far the histidine allele has been detected in only one cell line, HCT116 (7). The rare his-33 form of the protein is 3-4 fold more resistant to FdUrd than the tyr-33 form, due to an 8 fold lower catalytic efficiency (kcat), suggesting that histidine at residue 33 perturbs the structure of the TS active site (9)

*The human thymidylate synthase gene maps to chromosome 18p11.32*

5 The gene for human thymidylate synthase was initially mapped to the long arm of chromosome 18 (18q21.31-qter) by somatic cell hybrid analysis (10), however two subsequent reports place the gene in band 18p11.32 using fluorescence *in situ* hybridization (11,12).

*Chromosome band 18p11.32 is a site of loss of heterozygosity*

10 The long arm of chromosome 18 contains the DCC (deleted in colon cancer) candidate tumor suppressor gene and has been well studied in a variety of tumors. The short arm (18p), where TS apparently resides, has not been studied as extensively. The available data suggests there is LOH in approximately 45% of colon cancers (13) and 25-30% of cervical (14), head and neck (15), lung (16) and ovarian (17) cancers and sarcomas.  
15 LOH has also been described in breast, brain, esophagus, kidney and prostate cancers (0-15%). 18p has not been studied for allele loss in several other major cancers, including bladder, leukemia, lymphoma, liver, pancreas, stomach and testicular cancers.

20 **References**

1. Chu, E., Koeller, D.M., Casey, J.L., et al. (1991) Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase. *Proc. Natl. Acad. Sci. U.S.A.* 88: 8977-81.
- 25 2. Seno, T., Ayusawa, D., Shimizu, K., et al. (1985) in de Serres, F.J. (ed.) Genetic Consequences of Nucleotide Pool Imbalance, Plenum Publishing Company, New York, pp. 241-263.
3. Johnson, L.F. (1994) Posttranscriptional regulation of thymidylate synthase gene expression. *Journal of Cellular Biochemistry* 54: 387-392.

4. Calabresi, P. and B. Chabner (1996) in Hardman, J.G., Limbird, L.E., et al. (eds.) Goodman and Gilman's The Pharmacological Basis of Therapeutics, McGraw Hill, New York, pp. 1247-1251.

5. Takeishi, K., Kaneda, S., Aysawa, D., Shimizu, K., Gotoh, O. and T. Seno (1985) Nucleotide sequence of a functional cDNA for human thymidylate synthase. *Nucleic Acids Research* 13: 2035-2043.

6. Kaneda, S., Nalbantoglu, K., Takeishi, K., et al. (1990) Structural and functional analysis of the human thymidylate synthase gene. *Journal of Biological Chemistry* 265: 20277-84.

10 7. Barbour, K.W., Berger, S.H. and S.G. Berger (1990) Single amino acid substitution defines a naturally occurring genetic variant of human thymidylate synthase. *Molecular Pharmacology* 37: 515-518.

15 8. Barbour, K.W., Hoganson, D.K., Berger, S.H. and F.G. Berger (1992) A naturally occurring tyrosine to histidine replacement at residue 33 of human thymidylate synthase confers resistance to 5-fluoro-2'-deoxyuridine in mammalian and bacterial cells. *Molecular Pharmacology* 42: 242-248

9. Hughey, C.T., Barbour, K.W., Berger, F.G. and S.H. Berger (1993) Functional effects of a naturally occurring amino acid substitution in human thymidylate synthase. *Molecular Pharmacology* 44: 316-323.

20 10. Nussbaum, R.L., McCarrick-Walmsley, R., Lesko, J.G., et al. (1985) Thymidylate synthase deficient Chinese hamster cells: a selection system for human chromosome 18 and experimental system for the study of thymidylate synthase regulation and fragile X expression. *American Journal of Human Genetics* 37: 1192-1205.

25 11. Hori, T., Takahashi, E., Ayusawa, D., et al. (1990) Regional assignment of the human thymidylate synthase gene to chromosome band 18p11.32 by nonisotopic *in situ* hybridization. *Human Genetics* 85: 576-580.

12. Silverman, G.A., Kuo, W.-L., Taillon-Miller, P. and J.W. Gray (1993) Chromosomal reassignment: YACs containing both YES1 and thymidylate synthase map to the short arm of chromosome 18. *Genomics* 15: 442-445.

13. Vogelstein, B., Fearon, E.R., Kern, S.E., et al. (1989) Allelotype of colorectal carcinomas. *Science* 244: 207-211.
14. Mullokandov, M.R., Kholodilov, N.G., Atkin, N.B., et al. (1996) Genomic Alterations in cervical carcinoma: losses of chromosome heterozygosity and human papilloma virus tumor status. *Cancer Research* 56: 197-205.
15. Nawroz, H., van der Riet, P., Hruban, R.H., et al. (1994) Allelotype of head and neck squamous cell carcinoma. *Cancer Research* 54: 1152-55.
16. Allelotype of non-small cell lung carcinoma - comparison between loss of heterozygosity in squamous cell carcinoma and adenocarcinoma. *Cancer Research*: 52: 2478-81.
17. Abeln, E.C.A., Kuipers-Dijkshoorn, N.J., Berns, E.M.J.J., et al. (1995) Molecular genetic evidence for unifocal origin of advanced epithelial ovarian cancer and for minor clonal divergence. *British Journal of Cancer* 72: 1330-1336.

#### **Example 14: Cytidine Triphosphate Synthetase (CTPS) - Target Gene VARIA260**

*Cytidine Triphosphate Synthetase is essential for cell growth*

Human cytidine triphosphate synthetase catalyzes the glutamination of UTP to form CTP. The reaction is:  $UTP + ATP + \text{glutamine} \rightarrow CTP + ADP + P_i + \text{glutamate}$ . This is the rate limiting step in the synthesis of cytidine nucleotides from both the *de novo* and uridine salvage synthesis routes (see ref. 1 and references therein). CTPS also plays a vital regulatory function in balancing nucleotide pools for DNA polymer synthesis; it is allosterically regulated by CTP (negatively) and GTP (positively).

There is compelling evidence that CTPS is essential for cell survival:

CTPS is evolutionarily conserved in yeast and bacteria, with a high degree of amino acid identity in regions mediating allosteric regulation and catalysis (1-

212

232/116

3). (Another example: the human and hamster enzymes are identical in length and 98% amino acid identical over 591 amino acids.)

Mutant hamster cells lacking functional CTPS need exogenous cytidine to survive (3).

5 There is no known human deficiency disease of CTPS.

CTPS function is increased in proliferating cells (4).

10 Thus CTPS is an attractive target for antiproliferative drugs. Cyclopentyl cytosine (CPE-C) is a synthetic cytidine analog in which a cyclopentyl group replaces the furan ring of the ribose sugar. CPE-C has antineoplastic and antiviral effects in animal models (5). The drug is kinased intracellularly to the triphosphorylated nucleotide form (CPE-CTP). Exposure of cells to CPE-C leads to rapid depletion of CTP pools, as a result of inhibition of CTPS by CPE-CTP (6,7). Upregulation of CTP synthetase, or loss of negative allosteric modulation by CTP is associated with resistance to the cancer chemotherapy drugs arabinosyl cytosine (ara-C), 5-fluorouracil and other cytotoxic nucleoside analogs as well as alkylating agents (3).

15

*The human cytidine triphosphate synthetase gene has sequence variances*

20 The sequence of a human cytidine triphosphate synthetase cDNA was determined by Yamauchi et al. (1), who later determined the genomic sequence as well (2). We undertook a systematic search for DNA sequence variance by analysing 36 unrelated individuals using the single strand conformation polymorphism technique. Primers were designed using the sequence of Yamauchi et al. (1). SSCP analysis revealed 3

25 DNA fragments having sequence variances, and subsequent DNA sequence analysis showed that nucleotides 576 (A vs. G), 2093 (C vs. T) and 2135 (G vs. A) vary among normal individuals as shown in the Target Summary Table. The nucleotide 576 sequence variance is a silent substitution in the coding region, while the latter two sequence variances are in the 3' untranslated region of the cDNA. All three sequence

variances were detected at low frequency in the panel of 36 individuals (3-8%), however all but one of the heterozygotes is Asian, and it seems likely that a larger survey of Asian populations would show higher allele frequencies in Chinese and other groups. For example among the four Chinese in the panel two (50%) are heterozygous for the residue 2135 sequence variance, and one (25%) is heterozygous for the nt 576 sequence variance. Also, the one Cambodian in the panel is heterozygous for both the 2093 and 2135 sequence variances.

*The human cytidine triphosphate synthetase gene maps to chromosome 1p34.1*

The gene for human cytidine triphosphate synthetase has been mapped to 1p34.1 by somatic cell hybrid analysis (2).

*Chromosome band 1p34.1 is a site of frequent loss of heterozygosity*

The short arm of chromosome 1 is comparatively well investigated for allele loss, especially in breast and colon cancers. The 1p35-32 and 1p22-13 regions flank 1p34.1 and are the best available markers for LOH on 1p. Studies of these regions show 30-50% LOH frequency in breast cancer (8-12), 41-75% in glioma (a brain cancer subtype) (13), 20-40% in colon cancer (14,15), ~50% in stomach cancer (16), ~20% in lung cancer (17) and 20-30% in ovarian cancer (18). High frequency LOH has been detected in several uncommon cancers such as pheochromocytoma (50-86%) and neuroblastoma (~50%). Most other common cancers have not been adequately investigated to assess LOH frequency in this region.

## References

1. Yamauchi, M., Yamauchi, N. and M. Meuth (1990) Molecular cloning of the human CTP synthetase gene by functional complementation with purified human metaphase



chromosomes. *EMBO Journal* 9: 2095-2099.

2. Yamauchi, M., Yamauchi, N., Phear, G., et al. (1991) Genomic organization and chromo-somal localization of the human CTP synthetase gene(CTPS). *Genomics* 11: 1088-96.

5 3. Whelan, J., Phear, G., Yamauchi, M. and M. Meuth (1993) Clustered base substitutions in CTP synthetase conferring drug resistance in Chinese hamster ovary cells. *Nature Genetics* 3: 317- 322.

10 4. van den Berg, A., van Lenthe, H., Busch, S., et al. (1993) Evidence for transformation related increase in CTP synthetase activity *in situ* in human lymphoblastic leukemia. *European Journal of Biochemistry* 216: 161-167.

5. Marquez, V.E., Lim, M.-I., Treanor, S.P., et al. (1988) Cyclopentylcytosine: a carbocyclic nucleoside with antitumor and antiviral properties. *Journal of Medical Chemistry* 31: 1687-94.

15 6. Kang, G.J., Cooney, D.A., Moyer, J.D., et al. (1989) Cyclopentenyl triphosphate: formation and inhibition of CTP synthetase. *Journal of Biological Chemistry* 264: 713-718.

7. Glazer, R.I., Knode, M.C. Lim, M.-I., and V.E. Marquez (1985) Cyclopentyl cytidine analogue: an inhibitor of cytidine triphosphate synthesis in human colon carcinoma cells. *Biochemical Pharmacology* 34: 2535-2539.

20 8. Bieche I, Champeme MH, Matifas F, Cropp CS, Callahan R, Lidereau R. (1993) Two distinct regions involved in 1p deletion in human primary breast cancer. *Cancer Res.* 53:1990-4.

25 9. Borg A, Zhang QX, Olsson H, Wenngren E. (1992) Chromosome 1 alterations in breast cancer: allelic loss on 1p and 1q is related to lymphogenic metastases and poor prognosis. *Genes Chromosomes & Cancer.* 5:311-20.

10. Sato T, Tanigami A, Yamakawa K, et al. (1990) Allelotype of breast cancer: cumulative allele losses promote tumor progression in primary breast cancer. *Cancer Res.* 50:7184-9.

11. Devilee P, van Vliet M, Bardoel A, et al. (1991) Frequent somatic imbalance of

marker alleles for chromosome 1 in human primary breast carcinoma. *Cancer Res.* 51:1020-5.

12. Loupart ML, Armour J, Walker R, Adams S, Brammar W, Varley J. (1995) Allelic imbalance on chromosome 1 in human breast cancer. I. Minisatellite and RFLP analysis. *Genes Chromosomes & Cancer.* 12:16-23.

13. Reifenberger, J., Reifenberger, G., Liu, L., et al. (1994) Molecular genetic analysis of oligodendroglial tumors shows preferential allelic deletions on 19q and 1p. *American Journal of Pathology* 145: 1175-1190.

14 Meling GI, Lothe RA, Borresen AL, et al. (1991) Genetic alterations within the retinoblastoma locus in colorectal carcinomas. Relation to DNA ploidy pattern studied by flow cytometric analysis. *Br J Cancer.* 64:475-80.

15. Lothe RA, Nakamura Y, Woodward S, Gedde DT, Jr., White R. (1988) VNTR (variable number of tandem repeats) markers show loss of chromosome 17p sequences in human colorectal carcinomas. *Cytogenet Cell Genet.* 48:167-9.

16. Ezaki, T., Yanagisawa, A., Ohta, K., et al. ((1996) Deletion mapping on chromosome 1p in well-differentiated gastric cancer. *British Journal of Cancer* 73: 424-428.

17. Hiyama K, Ishioka S, Shirotani Y, et al. (1995) Alterations in telomeric repeat length in lung cancer are associated with loss of heterozygosity in p53 and Rb. *Oncogene.* 10:937-44.

18. Yang-Feng TL, Han H, Chen KC, et al. (1993) Allelic loss in ovarian cancer. *Int J Cancer.* 54:546-51.

#### **Example 15: Cysteinyl tRNA Synthetase (CARS) - Target Gene VARIA301**

*The human cysteinyl tRNA synthetase gene is essential for cell survival*

Cysteinyl-tRNA synthetase (CARS) catalyzes ATP dependent covalent attachment of

216

232/116

cysteine to its cognate tRNA to form cysteinyl-tRNA. In the absence of cysteinyl-tRNA, protein synthesis is blocked. Since Cysteinyl-tRNA synthetase is a single copy gene in man, inhibition of its function is expected to be cell lethal. This has been shown for other tRNA synthetases (summarized above).

5

*The human cysteinyl-tRNA synthetase gene and mRNA have sequence variances*

A human cDNA encoding cysteinyl tRNA synthetase (CARS) was cloned based on the similarity of a human expressed sequence tag (EST) to *E. coli* cysteinyl tRNA synthetase (1). The published human CARS cDNA is 2048 nucleotides in length and includes a 30 nucleotide 5' untranslated region followed by an open reading frame of 1914 nucleotides and a 3' untranslated region of 134 nucleotides (1). An EMBL/GENBANK submission (accession # L06845) by the authors of ref. 1 includes a 3' untranslated region 423 nucleotides longer than the published sequence, but lacks 19 consecutive A nucleotides after position 2029 (making a net increase of:  $423 - 19 = 404$  nucleotides, and a composite cDNA of:  $2048 + 404 = 2452$  nucleotides in length. We have confirmed the existence of 2452 nt transcripts by PCR amplification of reverse transcribed mRNA.) We designed primers as shown on the annotated cDNA sequence and screened the composite 2452 nt cDNA for sequence variance in 36 unrelated individuals by the single strand conformation polymorphism (SSCP) technique. Two sequence variances were identified. One of the sequence variances, located in the 5' untranslated region, was below the desired level of 20% heterozygosity. The other sequence variance is a C vs. T transition near the 3' end of the coding sequence at nucleotide 1739 (see annotated sequence).

25

*The human cysteinyl tRNA synthetase protein has sequence variances*

The deduced amino acid sequence of the human CARS gene encodes a protein of 638 amino acids which probably functions as a monomer, by analogy to related synthetases. The deduced protein contains two sequence motifs, HIGH (residues 64-

217

232/116

67) and KMSKS (residues 406-410), which define Class I synthetases (see ref. 2 for background information on tRNA synthetases). These two conserved motifs form an ATP binding fold (the Rossman fold) in the amino terminal half of the protein. Cytosine at nucleotide 1739 encodes proline at residue 622 of the protein, while  
5 thymine at nucleotide 1739 encodes leucine. The pro/leu amino acid sequence variance is a mere 16 residues from the C terminus of the protein. The C-terminal portion of CARS, by analogy to other class I synthetases, contains the tRNA binding site.

10 *Frequency of CARS heterozygotes*

The frequency of heterozygotes for the nucleotide 1739 sequence variance is ~45-50% in all major racial groups surveyed (see accompanying table), including North American Whites (8/15=53%), North American Blacks (2/4=50%), Chinese  
15 (2/4=50%), Swedish (127/344=37%) and Japanese (1/4=25%). The wide population distribution of both alleles suggests that other population groups will also have a high frequency of heterozygotes.

20 *Gene Mapping of CARS to 11p15.5*

Human CARS cDNA has been mapped to chromosome 11p15.5 by screening human X Chinese hamster somatic cell hybrids informative for all human chromosomes, and by fluorescence *in situ* hybridization (3). Both mapping techniques were conclusive in showing only one locus for human CARS. Detailed physical maps of 11p15.5 have  
25 subsequently allowed precise localization of the CARS gene relative to other DNA markers (4).

*LOH at 11p15.5 is well documented in many cancer types*

The short arm of chromosome 11, and particularly the 11p15.5 region, is deleted in a

variety of human cancers, including (but not limited to) ovarian (18 - 50% LOH), non-small cell lung (22 - 71%), breast (12 - 33%), bladder (40 - 50%), esophageal (18 - 40%) and testicular cancers (18 - 66%) (refs. 5-12). Many deletions in the 11p15.5 region span relatively short chromosomal segments (2 - 10 megabases; see ref. 8).  
5 Using the specific variances identified in the CARS gene as markers for heterozygosity, we have determined that LOH occurs in 10/20 ovarian cancers (50%) and 10/52 non-small cell lung cancers (19%).

#### *Assays for human CARS inhibitors*

10 There is no published work on the protein encoded by the putative human CARS cDNA, nor on any other eukaryotic CARS protein, however the extensive characterization of other Class I synthetases from both prokaryotes and eukaryotes provides a template for modeling the structure of human CARS. (For an example of  
15 how this can be done see ref. 14, in which the three dimensional structure of human alanyl-tRNA synthetase has been modeled up to amino 249 by neural net software and multiple alignments of partial and complete human AARS sequences with heterologous prokaryotic class II synthetases for which crystal structures exist.) With  
20 respect to the C-terminal location of the variant amino acid residue in human CARS, it is worth noting that single amino acid substitutions in the C-terminal region of alanyl tRNA synthetase can have greater than 100 fold effects on catalytic activity (15).

#### **References**

- 25 1. Wasmuth, J.J. Cruzen, M. E. and S.M. Arfin (1994) Nucleotide and deduced amino acid sequence of human cysteinyl-tRNA synthetase. *DNA Sequence* 4:243-248.  
2. Moras, D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends in Biochemical Sciences* 17: 159-164.  
3. Cruzen, M.E., Bengtsson, U., McMahon, J., Wasmuth, J.J. and S.M. Arfin (1993)

Assignment of the cysteinyl-tRNA synthetase gene (CARS) to 11p15.5. *Genomics* 15: 692-693.

- 5        5. Winqvist, R., Mannermaa, A., Alavaikko, M., Blanco, G., Taskinen, P.J., Kiviniemi, H., Newsham, I. and W. Cavenee (1993) Refinement of regional loss of heterozygosity for chromosome 11p15.5 in human breast tumors. *Cancer Research* 53: 4486-4488.
6. Kiechle-Schwartz, M., Bauknecht, T., Wienker, T., et al. (1993) Loss of Constitutional Heterozygosity on Chromosome 11p in Human Ovarian Cancer. *Cancer* 72:2423-32.
- 10       7. Viel, A., Giannini, F., Tumiotti, L., Sopracordevole, F., Visentin, M.C. and M. Boiocchi (1992) Chromosomal localisation of two putative 11p oncosuppressor genes involved in human ovarian tumors *British Journal of Cancer* 66: 1030-1036.
8. Bepler, G. and Garcia-Blanco, M.A. (1994) Three Tumor Suppressor Regions on Chromosome 11p Identified by High Resolution Deletion Mapping in Human Non-  
15       Small Cell Lung Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 91:5513-7.
9. Iizuka, M., Sugiyama, Y., Shiraishi, M., Jones, C. and T. Sekiya (1995) Allelic losses in human chromosome 11 in lung cancers. *Genes, Chromosomes & Cancer* 13:40-46.
10. Shaw, M.E. and Knowles, M.A. (1995) Deletion Mapping of Chromosome 11 in Carcinoma of the Bladder. *Genes, Chromosomes & Cancer* 13:1-8.
- 20       11. Smith, R.C., and Rukstalis, D.B. Frequent Loss of Heterozygosity at 11p Loci in Testicular Cancer. *The Journal of Urology* 153:1684-7, 1995.
12. Shibagaki, I., Shimada, Y., Wagata, T., Ikenaga, M., Imamura, M. and K. Ishizaki (1994) Allelotype analysis of esophageal squamous cell carcinoma. *Cancer Research* 54: 2996-3000.
- 25       13. Shiba, K., Suzuki, N., Shigesada, K., Namba, Y., Schimmel, P. and T. Noda (1994) Human cytoplasmic isoleucyl-tRNA synthetase: selective divergence of the anticodon-binding domain and acquisition of a new structural unit. *Proc. Natl. Acad. Sci. U.S.A.* 91:7435-7439.
14. Shiba, K., Ripmaster, T., Suzuki, N., Nichols, R., Plotz, P., Noda, T. and P.

220

232/116

Schimmel (1995) Human alanyl-tRNA synthetase: conservation in evolution of catalytic core and microhelix recognition. *Biochemistry* 34: 10340-10349.

15. Wu, M.-X., Filley, S.J., Xiong, J., Lee, J.J. and K.A.W. Hill (1994) A cysteine in the C-terminal region of alanyl-tRNA synthetase is important for aminoacylation activity. *Biochemistry* 33: 12260-12266.

5

**Example 16: Glutamyl-Prolyl tRNA Synthetase (EPRS): - Target Gene VARIA300**

10

*The human glutamyl-prolyl tRNA synthetase gene is essential for cell survival*

Glutamyl-prolyl-tRNA synthetase (EPRS) catalyzes ATP dependent covalent attachment of glutamine and proline to their cognate tRNAs to form glutamyl-tRNA and prolyl-tRNA. In the absence of glutamyl-tRNA or prolyl-tRNA, protein synthesis is blocked. Since glutamyl-prolyl-tRNA synthetase is a single copy gene in man, inhibition of its function is expected to be cell lethal. This has been shown for other tRNA synthetases (summarized above).

15

20

*The human glutamyl-prolyl tRNA synthetase gene, mRNA and protein have sequence variances*

A human cDNA encoding glutamyl-prolyl tRNA synthetase (EPRS) was initially misidentified as glutaminyl-tRNA synthetase (1) based on misleading sequence alignments with bacterial and yeast glutaminyl-tRNA synthetase (2). Subsequently, biochemical studies of the protein encoded by a *D. melanogaster* gene ~70% identical to the human gene demonstrated glutamyl (not glutaminyl) tRNA synthetase activity, and also showed that a single gene encodes both glutamyl- and prolyl-tRNA synthetases in the fly (3). These observations eventually led to the realization that

25

human EPRS is also a single polypeptide containing two synthetases (2). The aminoacyl tRNA synthetases are divided into two classes (see *Background on tRNA Synthetases*, above). Glutamyl-tRNA synthetase belongs to Class I while Prolyl-tRNA synthetase belongs to class II. Thus the two halves of EPRS evolved independently and likely represent an evolutionarily recent fusion. The published human EPRS cDNA is 4,586 nt long and includes a 5' untranslated region of 58 nt followed by an open reading frame of 4320 nt and a 3' untranslated sequence of 208 nt (1). The gene encodes a polypeptide of 1440 amino acids. The glutamyl-tRNA synthetase activity is encoded by an imprecisely defined segment at 5' end of the gene probably spanning at least amino acids 105-426, while the prolyl-tRNA synthetase activity is encoded by a segment likely including residues 942-1369 at the 3' end of the gene (2). The two synthetase moieties are connected by a central domain of unknown function. It has been speculated that the central domain may attach the enzyme to the cytoskeleton or to other aminoacyl-tRNA synthetases in a multienzyme complex (2, 3).

*The human glutamyl-prolyl-tRNA synthetase gene and mRNA have sequence variances*

We designed primers and screened the 4586 nt cDNA for sequence variance in 36 unrelated individuals by the single strand conformation polymorphism technique. Seven sequence variances were identified, four located in the coding sequence and three located in the 3' untranslated region. As shown on the Annotated Glutamyl-Prolyl tRNA Synthetase cDNA Sequence and in the Target Summary Page, the sequence variance nucleotides are 2520 (C vs. A), 2944 (G vs. A), 2963 (C vs. T), 2969 (A vs. G), 3247 (A vs. G), 4459 (G vs. A) and 4506 (G vs. A). The sequences flanking the alternate allelic forms and their frequencies of occurrence are shown on the Target Summary Page. Less than 10% of individuals surveyed are heterozygous for sequence variances at 2520, 2944 and 2963. Heterozygotes for the other 4 sequence variances occur more frequently and appear to be widely distributed in the surveyed populations (see below).



222

232/116

*The human glutamyl-prolyl tRNA synthetase protein has sequence variances* Three nucleotide sequence variances, at 2520, 2963 and 2969, alter the amino acid coding sequence of EPRS at residues 821 (pro/his), 969 (his/tyr) and 971 (ile/val). The residue 821 his and 969 tyr alleles are relatively rare, with fewer than 10% heterozygotes in the surveyed populations. The more common residue 971 sequence variance lies in the PRS domain of the protein, near one of the widely conserved defining motifs for class II tRNA synthetases.

*EPRS heterozygotes are frequent in non-Asian populations.* While the overall frequency of residue 971 heterozygotes is 8/36 (24%), the frequency of heterozygotes varies among different populations. For example, there are no heterozygotes among 10 Asians surveyed (Chinese, Japanese, Filipino and Korean), while 8/26 (31%) of non-Asians, including North American Whites, Blacks and Hispanics, are heterozygotes.

#### *The EPRS Gene Maps to 1q41-q42*

Human EPRS cDNA has been mapped to chromosome 1q41-42 by screening human X Chinese hamster somatic cell hybrids informative for all human chromosomes, and by fluorescence *in situ* hybridization (3). Both mapping techniques were conclusive in showing only one locus for human EPRS.

*Loss of heterozygosity at 1q41-42 is documented in several cancer types.* 17-25% of breast cancers have allele loss in the 1q41-q42 region (4, 5), 29-46% of colon cancers (6, 7) and 17-26% of cervical cancers (8). One report describes 27% LOH in stomach cancer (9). One or two studies of brain, esophageal, kidney, liver and ovarian cancers also report LOH. No studies of LOH in the 1q41-q42 region have been reported in bladder, endocrine, head and neck, lung, or pancreas cancers or in leukemia or lymphoma.

*Antisense considerations* The sequence variances at 2963 and 2969 are close enough that a 20-mer antisense oligonucleotide could easily span them. Such an oligonucleotide should afford greater allele discrimination than is possible with a single nucleotide difference. However, the 2963 sequence variance is fairly rare (<10% heterozygotes) and not in linkage disequilibrium with the 2963 sequence variance, so there are more than two haplotypes in the populations tested.

### References

1. Fett, R. and R. Knippers (1991) The primary structure of human glutamyl tRNA synthetase. *Journal of Biological Chemistry* 266: 1448-1455.
2. Cerini, C., Kerjan, P., Astier, M., Gratecos, D., Mirande, M. and M. Semeriva (1991) A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase. *The EMBO Journal* 10: 4267-4277.
3. Kaiser, E., Hu, B., Becher, S., Eberhard, D., et al. (1994) The human EPRS locus (formerly the QARS locus): a gene encoding a class I and a class II aminoacyl-tRNA synthetase. *Genomics* 19: 280-290.
4. *Journal of The National Cancer Institute* 84: 506.
5. *Cancer Research* 51: 1020.
6. *International Journal of Cancer* 53: 382.
7. *Genes, Chromosomes & Cancer* 12: 16.
8. *Cancer Research* 56: 197.
9. *Cancer Research* 52: 3099.
10. Shiba, K., Ripmaster, T., Suzuki, N., Nichols, R., Plotz, P., Noda, T. and P. Schimmel (1995) Human alanyl-tRNA synthetase: conservation in evolution of catalytic core and microhelix recognition. *Biochemistry* 34: 10340-10349.

**Example 17: Alanyl-tRNA Synthetase (AARS) - Target Gene VARIA304**

*The human glutamyl-prolyl tRNA synthetase gene is essential for cell survival*

Alanyl-tRNA synthetase (AARS) catalyzes ATP dependent covalent attachment of alanine to its cognate tRNA to form alanyl-tRNA. In the absence of alanyl-tRNA,  
5 protein synthesis is blocked. Since alanyl-tRNA synthetase is a single copy gene in man (see below) inhibition of its function is expected to be cell lethal. This has been shown for other tRNA synthetases (summarized above).

*The human alanyl-tRNA synthetase gene and mRNA have sequence variances*

10

A human cDNA encoding alanyl tRNA synthetase (AARS) was cloned by Shiba et al. (1) using cross species PCR: AARS sequences from four evolutionarily distant species were compared and primers were designed to conserved regions specific to AARS. The cloned human cDNA is 3344 nt in length and includes a 110 nt 5' untranslated  
15 region, an open reading frame of 2904 nt encoding a 968 residue polypeptide, and a 3' untranslated region of 330 nt (ref. 1; Genbank accession D32050).

20

We designed primers. The 3344 nt cDNA was screened for sequence variance in 36 unrelated individuals by the single strand conformation polymorphism (SSCP) technique. One sequence variance was identified, a C vs. T transition at nucleotide 1013, within the coding sequence. The published nucleotide at position 1013 is T (1).

25

*The frequency of AARS heterozygotes is 25-50% in all populations surveyed.* The frequency of heterozygotes for the nucleotide 1013 sequence variance is 57% in the 36 individuals tested. Both alleles are present in all major racial groups surveyed (see Target Gene Summary Table), including North American Whites (9/15=60% heterozygotes), North American Blacks (3/4=75%), Chinese (2/4=50%), Japanese (1/4=25%) and Hispanic (1/2). The wide population distribution of both alleles suggests that other population groups will also have a high frequency of heterozygotes.

*The AARS gene maps to 16q22*

The human AARS cDNA has been mapped to chromosome 16q22 by us and by Nichols et al. (ref. 2). We designed primers to the 3' untranslated region of AARS and used PCR to analyze the National Institute of General Medical Sciences (NIGMS) Human/Rodent Somatic Cell Hybrid Mapping Panel #2 (see page 704 of the NIGMS 1994/1995 Catalog of Cell Lines, available from the Coriell Cell Repository, Camden, NJ). The panel consists of 24 hybrid cell lines, each monochromosomal for one human chromosome. The AARS PCR product mapped to the hybrid containing human chromosome 16. Subsequently we screened the Radiation Hybrid Mapping Panel created at Stanford University (rhserver@shgc.stanford.edu) and distributed by Research Genetics (RH01). The AARS PCR product mapped near D16S496 with a lod score >10. D16S496 is a polymorphic DNA marker at 16q22. The AARS PCR product mapped near D16S496 with a LOD score >10. DH16S496 is a polymorphic DNA marker at 16q22. (See, ref. 29 for a full explanation of modification hybrid mapping.) Similar results were obtained by Nichols et al., who mapped AARS by analysis of the same NIGMS hybrid mapping panel, by PCR mapping in a chromosome 16 regional mapping panel and by fluorescence *in situ* hybridization to metaphase chromosomes. All mapping techniques were conclusive in showing only one locus for human AARS.

*LOH at 16q22 is well documented in many cancer types.* Loss of heterozygosity studies of chromosome 16q have principally focused on breast and liver cancers. In six detailed studies of breast cancer in the 16q22 region LOH frequencies of 40-60% have been reported (refs 3-8). 16q22 LOH has been reported in 25-90% of liver cancers (9-13), with the average around 45%. Less extensive studies of other cancer types report 16q22 LOH in 19% of bladder cancers, 20% of colon cancers (14), 19-27% of esophageal cancers (15), 25% of small cell lung cancers (16), 16-37% of ovarian cancers (17-19) and 22% of uterine cancers (20), and 31-50% of prostate cancers (21-

22).

### References

1. Shiba, K., Ripmaster, T., Suzuki, N., Nichols, R., Plotz, P., Noda, T. and P. Schimmel (1995) Human alanyl-tRNA synthetase: conservation in evolution of catalytic core and microhelix recognition. *Biochemistry* 34: 10340-10349.
2. Nichols, R.C., Pai, S.I., Ge, Q., Targoff, I.N., Plotz, P.H. and P. Liu (1995) Localization of two human autoantigen genes by PCR screening and *in situ* hybridization - glycyl tRNA synthetase locates to 7p15 and alanyl-tRNA synthetase locates to 16q22. *Genomics* 30:131-132.
3. Cleton-Jansen AM, Moerland EW, Kuipers-Dijkshoorn NJ, et al. (1994) At least two different regions are involved in allelic imbalance on chromosome arm 16q in breast cancer. *Genes Chromosom Cancer*. 9:101-7.
4. Dorion-Bonnet F, Mautalen S, Hostein I, Longy M. (1995) Allelic imbalance study of 16q in human primary breast carcinomas using microsatellite markers. *Genes Chromosomes Cancer*. 14:171-81.
5. Kashiwaba M, Tamura G, Suzuki Y, et al. (1995) Epithelial-cadherin gene is not mutated in ductal carcinomas of the breast. *Jpn J Cancer Res*. 86:1054-9.
6. O'Connell P, Pekkel V, Fuqua S, Osborne CK, Allred DC. (1994) Molecular genetic studies of early breast cancer evolution. *Breast Cancer Res Treat*. 32:5-12.
7. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res*. 51:5794-9.
8. Tsuda H, Callen DF, Fukutomi T, Nakamura Y, Hirohashi S. (1994) Allele loss on chromosome 16q24.2-qter occurs frequently in breast cancers irrespectively of differences in phenotype and extent of spread. *Cancer Res*. 54:513-7.
9. Fujimori M, Tokino T, Hino O, et al. (1991) Allelotype study of primary hepatocellular carcinoma. *Cancer Res*. 51:89-93.
10. Fujimoto Y, Hampton LL, Wirth PJ, Wang NJ, Xie JP, Thorgeirsson SS. (1994) Alterations of tumor suppressor genes and allelic losses in human hepatocellular

carcinomas in China [see comments]. *Cancer Res.* 54:281-5.

11. Tsuda H, Zhang WD, Shimosato Y, et al. (1990) Allele loss on chromosome 16 associated with progression of human hepatocellular carcinoma. *Proc Natl Acad Sci U S A.* 87:6791-4.

5 12. Tsuda H, Oda T, Sakamoto M, Hirohashi S. (1992) Different pattern of chromosomal allele loss in multiple hepatocellular carcinomas as evidence of their multifocal origin. *Cancer Res.* 52:1504-9.

13. Zhang WD, Hirohashi S, Tsuda H, et al. (1990) Frequent loss of heterozygosity on chromosomes 16 and 4 in human hepatocellular carcinoma. *Jpn J Cancer Res.* 81:108-11.

14. Ookawa K, Sakamoto M, Hirohashi S, et al. (1993) Concordant p53 and DCC alterations and allelic losses on chromosomes 13q and 14q associated with liver metastases of colorectal carcinoma. *Int J Cancer.* 53:382-7.

15. *Genes, Chromosomes & Cancer* 10: 177.

15 16. Yokota J, Wada M, Shimosato Y, Terada M, Sugimura T. (1987) Loss of heterozygosity on chromosomes 3, 13, and 17 in small-cell carcinoma and on chromosome 3 in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A.* 84:9252-6.

17. *Cancer Research* 51: 5118.

18. Osborne RJ, Leech V. (1994) Polymerase chain reaction allelotyping of human ovarian cancer. *Br J Cancer.* 69:429-38.

20 19. Yang-Feng TL, Han H, Chen KC, et al. (1993) Allelic loss in ovarian cancer. *Int J Cancer.* 54:546-51.

20. Okamoto A, Sameshima Y, Yamada Y, et al. (1991) Allelic loss on chromosome 17p and p53 mutations in human endometrial carcinoma of the uterus. *Cancer Res.* 51:5632-5.

25 21. Carter, B.S., Ewing, C.M., Ward, S.W., et al. (1990) Allelic loss of chromosomes 16q and 10q in human prostate cancer. *Proc Natl Acad Sci U S A.* 87: 8751-8755.

22. Bergerheim, U.S.R., Kunimi, K., Collins, V.P. and P. Ekman (1991) Deletion mapping of chromosomes 8, 10, and 16 in human prostatic carcinoma. *Genes,*

*Chromosomes & Cancer* 3: 215-220.

23. Boehnke, M., Lange, K. and D.R. Cox (1991) Statistical methods for multipoint radiation hybrid mapping. *Am. J. Hum. Genet.* 49: 1174-88.

5

**Example 18: Threonyl-tRNA Synthetase (TARS) - Target Gene VARIA302**

*The human threonyl-tRNA synthetase gene is essential for cell survival*

10 Threonyl-tRNA synthetase (TARS) catalyzes ATP dependent covalent attachment of threonine to its cognate tRNA to form threonyl-tRNA. In the absence of threonyl-tRNA, protein synthesis is blocked. Threonyl-tRNA synthetase is a single copy gene in man (see below) and inhibition of TARS is cell lethal. This has been shown using the specific TARS inhibitor borrelidin, a threonine analog. Borrelidin resistant CHO  
15 cell lines have been isolated; the most resistant lines contain ~60-100 fold more immunologically reactive protein and 10-20 fold higher TARS activity than non-selected CHO cells (1-3).

The human TARS enzyme is a homodimeric member of the class II tRNA synthetases.  
20 The human protein is 53% amino acid identical to *S. cerevisiae* cytoplasmic TARS, 40% amino acid identical to *E. coli* TARS and 39% amino acid identical to yeast mitochondrial TARS. The degree of evolutionary conservation is 52-64% when conservative substitutions are allowed.

25 *The human Threonyl-tRNA synthetase gene and mRNA have sequence variances.* A human cDNA encoding threonyl tRNA synthetase was cloned by Cruzen and Arfin (GENBANK accession M63180; ref. 2) using anti-TARS antibodies to screen a  $\lambda$ gt11 expression library. The cDNA is 2644 nt in length and includes a 138 nt 5' untranslated region, an open reading frame of 2136 nt encoding a 712 residue polypeptide, and a 3'

229

232/116

untranslated region of 370 nt.

We designed primers for amplification. The 2644 nt cDNA was screened for sequence variance in 36 unrelated individuals by the single strand conformation polymorphism (SSCP) technique. Three sequence variances were identified: G vs. A transitions at nucleotides 1608 and 1755 within the coding sequence, and a C vs. T transition at nucleotide 2395 of the 3' untranslated region. None of the sequence variances alters the sense of the coding strand. The published sequence shows G, G and T at the three sequence variance sites

*The frequency of TARS heterozygotes is 25-45% in all populations surveyed.* The nucleotide 1608 sequence variance was genotyped only in North American Whites, 45% of whom were heterozygotes. The nucleotide 1608 and 1755 sequence variances were both genotyped in 36 individuals, with overall heterozygosity rates of 31% and 25%, respectively. Both sequence variances were detected in North American Whites, North American Blacks, Hispanics and Chinese. Of 14 North American Whites genotyped at all 3 sequence variance nucleotides, 11 (79%) were heterozygous for a least one polymorphism (see threonyl tRNA synthetase summary table).

*The TARS gene maps to 5p13-CEN.* The human TARS cDNA has been mapped to chromosome 5p13-CEN by analysis of TARS isoelectric focusing patterns in human/Chinese hamster hybrids (). The mapping studies were consistent with one human TARS locus.

*LOH at 5p13-CEN is documented in several cancer types.* The best data on 5p LOH is in cervical cancer where 9 markers have been tested in 3 different studies. The frequency of LOH ranges from 12-57%, averaging ~45%. Other cancers that have been studied are breast (10-24% LOH), head and neck (20% LOH), adenocarcinoma of the lung (40% LOH, but only 5 cancers were studied), melanoma (40%) and ovary (15-



21%).

*Assays for human TARS inhibitors.* Human TARS protein is a homodimeric class II synthetase. Antibodies to rat TARS were used to clone the human protein. The high degree of amino acid conservation throughout the protein suggests that it may be possible to create yeast and/or bacterial strains with human CARS.

### References

1. Gantt, J.S., Bennett, C.A. and S.M. Arfin (1981) Increased levels of threonyl tRNA synthetase in a borrelidin-resistant Chinese hamster ovary cell line. *Proc. Natl. Acad. Sci. U. S. A.* 92: 5367-5370.
2. Gerken, S.C. and S.M. Arfin (1984) Chinese hamster ovary cells resistant to borrelidin overproduce threonyl-tRNA synthetase. *The Journal of Biological Chemistry* 259: 9202-9206.
3. Kontis, K.J. and S.M. Arfin (1989) Isolation of a cDNA clone for human threonyl tRNA synthetase: amplification of the structural gene in borrelidin resistant cell lines. *Molecular and Cellular Biology* 9: 1832-1838.
4. Cruzen, M.E. and S.M. Arfin (1991) Nucleotide and deduced amino acid sequence of human threonyl-tRNA synthetase reveals extensive homology to the Escherichia coli and yeast enzymes. *The Journal of Biological Chemistry* 266: 9919-9923.
5. Gerken, S.C., Wasmuth, J.J. and S.M. Arfin (1986) Threonyl-tRNA synthesis gene maps close to leucyl-tRNA synthetase gene on human chromosome 5. *Somatic Cell and Molecular Genetics* 12: 519-522.
6. Mitra AB, Murty VV, Singh V, et al. (1995) Genetic alterations at 5p15: a potential marker for progression of precancerous lesions of the uterine cervix. *J Natl Cancer Inst.* 87:742-5.
7. Mitra AB, Murty VV, Li RG, Pratap M, Luthra UK, Chaganti RS. (1994) Allelotype analysis of cervical carcinoma. *Cancer Res.* 54:4481-7.
8. Mullokandov MR, Kholodilov NG, Atkin NB, Burk RD, Johnson AB, Klinger HP.

(1996) Genomic alterations in cervical carcinoma: losses of chromosome heterozygosity and human papilloma virus tumor status. *Cancer Res.* 56:197-205.

9. Larsson C, Bystrom C, Skoog L, Rotstein S, Nordenskjold M. (1990) Genomic alterations in human breast carcinomas. *Genes Chromosomes Cancer.* 2:191-7.

5 10. *Cancer Research* 54:1152

11. Wieland I, Bohm M, Arden KC, et al. (1996) Allelic deletion mapping on chromosome 5 in human carcinomas. *Oncogene.* 12:97-102.

12. Dracopoli NC, Houghton AN, Old LJ. (1985) Loss of polymorphic restriction fragments in malignant melanoma: implications for tumor heterogeneity. *Proc Natl Acad Sci U S A.* 82:1470-4.

13. Osborne RJ, Leech V. (1994) Polymerase chain reaction allelotyping of human ovarian cancer. *Br J Cancer.* 69:429-38.

15 **Example 19: Glutaminyl-tRNA Synthetase (QARS) - Target Gene VARIA305**

*The human glutaminyl-tRNA synthetase gene is essential for cell survival*

20 Glutaminyl-tRNA synthetase (QARS) catalyzes ATP dependent covalent attachment of glutamine to its cognate tRNA to form glutaminyl-tRNA. In the absence of glutaminyl-tRNA, protein synthesis is blocked in eucaryotic cells. Glutaminyl-tRNA synthetase is a single copy gene in man. Inhibition of its function is expected to be cell lethal, as shown for other tRNA synthetases (summarized above).

25 *The human Glutaminyl-tRNA synthetase gene and mRNA have sequence variances.*

A human cDNA encoding glutaminyl tRNA synthetase (QARS) was cloned by Lamour et al. (1) who expressed the cDNA in *E. coli* and demonstrated glutaminyl tRNA synthetase activity in bacterial extracts. The cloned human cDNA

232

232/116

(Genbank/EMBL accession number X76013) is 2437 nt in length and includes a 5' untranslated region of 5 nucleotides, an open reading frame of 2325 nucleotides encoding a 775 amino acid polypeptide, and a 3' untranslated region of 107 nt including 8 terminal nt of poly A.

5

We designed primers for amplification. The QARS cDNA was screened for sequence variance in 36 unrelated individuals using the single strand conformation polymorphism (SSCP) technique. One sequence variance was identified, a C vs. T transition at nucleotide 404, within the coding sequence. The published nucleotide at position 404 is C. The sequence variance does not affect the protein encoded.

10

The frequency of heterozygotes for the nucleotide 404 sequence variance is 11% in the 36 individuals tested (4/36). However three of 16 North American Whites are heterozygotes (19%), and one of four Japanese (25%) (see Target Gene Summary Table).

15

*The QARS gene maps to 3p*

The human QARS cDNA has been mapped to chromosome 3 by hybridization of a QARS probe to a panel of 25 human/rodent somatic cell hybrids (1). One somatic cell hybrid, not known to contain human chromosome 3, was positive for both the QARS probe and an ACY1 probe. ACY1 maps to human 3p21, suggesting QARS may also map in this area. We independently mapped QARS to chromosome 3 using primers from the 3' untranslated region to analyze the National Institute of General Medical Sciences (NIGMS) Human/Rodent Somatic Cell Hybrid Mapping Panel #2 by PCR (see page 704 of the NIGMS 1994/1995 Catalog of Cell Lines, available from the Coriell Cell Repository, Camden, NJ). The panel consists of 24 hybrid cell lines, each monochromosomal for one human chromosome. The QARS PCR product mapped to the hybrid containing human chromosome 3. All mapping techniques were conclusive

20

25

in showing only one locus for human QARS.

*Chromosome band 3p21 is a site of frequent loss of heterozygosity.* The short arm of chromosome 3 has been well studied in breast, cervical, esophageal, kidney, and lung cancers. These studies report frequent allele loss at 3p21, varying up to 100% in some studies of small cell lung cancer. Among other cancers LOH occurs in approximately 20-30% of breast cancers (2,3), 30-60% of cervical cancers (4,5), 10-40% of esophageal cancers (6,7), 45-80% of kidney cancers (8-10), 50-100% of nasopharyngeal cancers (11), 0-75% of squamous cell head and neck cancers (12), 30-60% of melanomas (13), 30-100% of non-small cell lung cancers (14-16) and 80-100% in small cell lung cancer (17-19). Other for which there are reports of LOH in at least 20% of cases include leukemia, pancreas cancer, sarcoma, testis cancer and ovarian cancer. Other cancer types, including bladder and lymphoma, have not been studied for LOH at 3p21.

#### References

1. Nomura, N., Nagase, T., Miyajima, N., et al. (1994) Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Research* 1:225-229.
2. Nichols, R.C., Blinder, J., Pai, S.I. et al. (1996) Assignment of two human autoantigen genes: isoleucyl tRNA synthetase locates to 9q21 and lysyl-tRNA synthetase locates to 16q23-24. *Genomics*: 210-213.
3. Cleton-Jansen AM, Moerland EW, Kuipers-Dijkshoorn NJ, et al. (1994) At least two different regions are involved in allelic imbalance on chromosome arm 16q in breast cancer. *Genes Chromosom Cancer*. 9:101-7.
4. Dorion-Bonnet F, Mautalen S, Hostein I, Longy M. (1995) Allelic imbalance study of 16q in human primary breast carcinomas using microsatellite markers. *Genes Chromosomes Cancer*. 14:171-81.

5. Kashiwaba M, Tamura G, Suzuki Y, et al. (1995) Epithelial-cadherin gene is not mutated in ductal carcinomas of the breast. *Jpn J Cancer Res.* 86:1054-9.
6. O'Connell P, Pekkel V, Fuqua S, Osborne CK, Allred DC. (1994) Molecular genetic studies of early breast cancer evolution. *Breast Cancer Res Treat.* 32:5-12.
- 5 7. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* 51:5794-9.
8. Tsuda H, Callen DF, Fukutomi T, Nakamura Y, Hirohashi S. (1994) Allele loss on chromosome 16q24.2-qter occurs frequently in breast cancers irrespectively of differences in phenotype and extent of spread. *Cancer Res.* 54:513-7.
- 10 9. Fujimori M, Tokino T, Hino O, et al. (1991) Allelotype study of primary hepatocellular carcinoma. *Cancer Res.* 51:89-93.
10. Fujimoto Y, Hampton LL, Wirth PJ, Wang NJ, Xie JP, Thorgeirsson SS. (1994) Alterations of tumor suppressor genes and allelic losses in human hepatocellular carcinomas in China [see comments]. *Cancer Res.* 54:281-5.
- 15 11. Tsuda H, Zhang WD, et al. (1990) Allele loss on chromosome 16 associated with progression of human hepatocellular carcinoma. *Proc Natl Acad Sci U S A.* 87:6791-4.
12. Tsuda H, Oda T, Sakamoto M, Hirohashi S. (1992) Different pattern of chromosomal allele loss in multiple hepatocellular carcinomas as evidence of their multifocal origin. *Cancer Res.* 52:1504-9.
- 20 13. Zhang WD, Hirohashi S, Tsuda H, et al. (1990) Frequent loss of heterozygosity on chromosomes 16 and 4 in human hepatocellular carcinoma. *Jpn J Cancer Res.* 81:108-11.
14. Ookawa K, Sakamoto M, Hirohashi S, et al. (1993) Concordant p53 and DCC alterations and allelic losses on chromosomes 13q and 14q associated with liver metastases of colorectal carcinoma. *Int J Cancer.* 53:382-7.
- 25 15. *Genes, Chromosomes & Cancer* 10: 177.
16. *Cancer Research* 54: 2996.
17. Gallion H.H., Powell D.E., Morrow J.K., et al. (1992) Molecular genetic changes in human epithelial ovarian malignancies [see comments]. *Gynecol Oncol.* 47:137-42.

18. Osbome RJ, Leech V. (1994) Polymerase chain reaction allelotyping of human ovarian cancer. *Br J Cancer*. 69:429-38.
19. Yang-Feng TL, Han H, Chen KC, et al. (1993) Allelic loss in ovarian cancer. *Int J Cancer*. 54:546-51.
- 5 20. *British Journal of Urology* 73: 390.
21. Okamoto A, Sameshima Y, Yamada Y, et al. (1991) Allelic loss on chromosome 17p and p53 mutations in endometrial carcinoma of the uterus. *Cancer Res*. 51:5632-5.

10

**Example 20: Lysyl-tRNA Synthetase (KARS) - Target Gene VARIA303***Human Lysyl t-RNA synthase gene is essential*

15

Lysyl-tRNA synthetase (KARS) catalyzes ATP dependent covalent attachment of lysine to its cognate tRNA to form lysyl-tRNA. In the absence of lysyl-tRNA, protein synthesis is blocked. Since lysyl-tRNA synthetase is a single copy gene in man, inhibition of its function is expected to be cell lethal. This has been shown for other tRNA synthetases (summarized above).

20

*The human Lysyl-tRNA synthetase gene and mRNA have sequence variances*

25

A human cDNA encoding a sequence similar to bacterial lysyl tRNA synthetases was cloned by Nomura et al. (GenBank/DDBJ submission D31890; see ref. 1) while sequencing random cDNAs. No biochemical studies of the protein encoded by this sequence have been reported. The 5' end of the sequence apparently begins in the coding region and the open reading frame continues for 1805 nucleotides, encoding 601 residues of a polypeptide (the full length of which has not been established), followed by a 3' untranslated region of 165 nucleotides.

236

232/116

We designed primers for amplification. The reported partial cDNA was screened for sequence variance in 36 unrelated individuals using the single strand conformation polymorphism (SSCP) technique as described in the methods section. Two sequence variances were identified, an A vs. G transition at nucleotide 89 and a G vs. C transversion at nucleotide 1798, both within the coding sequence. The published nucleotides are A and G, respectively. The nucleotide 1798 sequence variance alters the sense of the 599th codon (the third codon from the end of the coding sequence) to serine vs. threonine.

*The frequency of KARS heterozygotes varies among the populations surveyed.* The frequency of heterozygotes for the nucleotide 89 sequence variance is 19% in the 36 individuals tested. However all heterozygous individuals were either North American Whites (4/16; 25% heterozygotes), North American Blacks (1/4; 25%), or Hispanics (1/3; 33% heterozygotes). The frequency of heterozygotes for the nucleotide 1798 sequence variance is 6% in the 36 individuals tested. However all heterozygous individuals were North American Blacks (2/4; 50%) (see Target Gene Summary Table). Further study of these and other population groups will better establish the frequency of heterozygotes for these two sequence variances.

*The KARS gene maps to 16q23-q24*

The human KARS cDNA has been mapped to chromosome 16q22 by Nichols et al. (ref. 2) and by us. We designed primers to the 3' untranslated region of KARS and used PCR to analyze the National Institute of General Medical Sciences (NIGMS) Human/Rodent Somatic Cell Hybrid Mapping Panel #2 (see page 704 of the NIGMS 1994/1995 Catalog of Cell Lines, available from the Coriell Cell Repository, Camden, NJ). The panel consists of 24 hybrid cell lines, each monochromosomal for one human chromosome. The KARS PCR product mapped to the hybrid containing human chromosome 16. Similar results were obtained by Nichols et al., who mapped KARS

237

232/116

by analysis of the same NIGMS hybrid mapping panel, by PCR mapping in a chromosome 16 regional mapping panel and by fluorescence *in situ* hybridization to metaphase chromosomes. The *in situ* hybridization showed KARS maps to 16q23-q24. All mapping techniques were conclusive in showing only one locus for human KARS.

*Loss of heterozygosity occurs frequently at 16q23-q24 in many cancer types.* Loss of heterozygosity studies of chromosome 16q have principally focused on breast and liver cancers. In six detailed studies of breast cancer in the 16q23-q24 region LOH frequencies of 30-60% have been reported (refs 3-8). 16q22 LOH has been reported in 35-65% of liver cancers (9-13), with the average around 45%. Studies of other cancer types report 16q22 LOH in 19% of colon cancers (14), 17-27% of esophageal cancers (15,16), 37% of ovarian cancers (new ref) (17-19), 18% of prostate cancers (20) and 23% of uterine cancers (21). Cancer types not yet investigated for LOH include kidney, leukemia and lymphoma, lung, melanoma, neuroblastoma, stomach and testis.

## References

1. Nomura, N., Nagase, T., Miyajima, N., et al. (1994) Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Research* 1:225-229.
2. Nichols, R.C., Blinder, J., Pai, S.I. et al. (1996) Assignment of two human autoantigen genes: isoleucyl tRNA synthetase locates to 9q21 and lysyl-tRNA synthetase locates to 16q23-24. *Genomics*: 210-213.
3. Cleton-Jansen AM, Moerland EW, Kuipers-Dijkshoorn NJ, et al. (1994) At least two different regions are involved in allelic imbalance on chromosome arm 16q in breast cancer. *Genes Chromosom Cancer*. 9:101-7.
4. Dorion-Bonnet F, Mautalen S, Hostein I, Longy M. (1995) Allelic imbalance study



of 16q in human primary breast carcinomas using microsatellite markers. *Genes Chromosomes Cancer*. 14:171-81.

5. Kashiwaba M, Tamura G, Suzuki Y, et al. (1995) Epithelial-cadherin gene is not mutated in ductal carcinomas of the breast. *Jpn J Cancer Res*. 86:1054-9.

5 6. O'Connell P, Pekkel V, Fuqua S, Osborne CK, Allred DC. (1994) Molecular genetic studies of early breast cancer evolution. *Breast Cancer Res Treat*. 32:5-12.

7. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res*. 51:5794-9.

10 8. Tsuda H, Callen DF, Fukutomi T, Nakamura Y, Hirohashi S. (1994) Allele loss on chromosome 16q24.2-qter occurs frequently in breast cancers irrespective of differences in phenotype and extent of spread. *Cancer Res*. 54:513-7.

9. Fujimori M, Tokino T, Hino O, et al. (1991) Allelotype study of primary hepatocellular carcinoma. *Cancer Res*. 51:89-93.

15 10. Fujimoto Y, Hampton LL, Wirth PJ, Wang NJ, Xie JP, Thorgeirsson SS. (1994) Alterations of tumor suppressor genes and allelic losses in human hepatocellular carcinomas in China [see comments]. *Cancer Res*. 54:281-5.

11. Tsuda H, Zhang WD, Shimosato Y, et al. (1990) Allele loss on chromosome 16 associated with progression of human hepatocellular carcinoma. *Proc Natl Acad Sci U S A*. 87:6791-4.

20 12. Tsuda H, Oda T, Sakamoto M, Hirohashi S. (1992) Different pattern of chromosomal allele loss in multiple hepatocellular carcinomas as evidence of their multifocal origin. *Cancer Res*. 52:1504-9.

13. Zhang WD, Hirohashi S, Tsuda H, et al. (1990) Frequent loss of heterozygosity on chromosomes 16 and 4 in human hepatocellular carcinoma. *Jpn J Cancer Res*. 81:108-11.

25 14. Ookawa K, Sakamoto M, Hirohashi S, et al. (1993) Concordant p53 and DCC alterations and allelic losses on chromosomes 13q and 14q associated with liver metastases of colorectal carcinoma. *Int J Cancer*. 53:382-7.

15. *Genes, Chromosomes & Cancer* 10: 177-

16. *Cancer Research* 54: 2996-
17. Gallion HH, Powell DE, Morrow JK, et al. (1992) Molecular genetic changes in human epithelial ovarian malignancies [see comments]. *Gynecol Oncol.* 47:137-42.
18. Osborne RJ, Leech V. (1994) Polymerase chain reaction allelotyping of human ovarian cancer. *Br J Cancer.* 69:429-38.
19. Yang-Feng TL, Han H, Chen KC, et al. (1993) Allelic loss in ovarian cancer. *Int J Cancer.* 54:546-51.
20. *British Journal of Urology* 73: 390-.
21. Okamoto A, Sameshima Y, Yamada Y, et al. (1991) Allelic loss on chromosome 17p and p53 mutations in human endometrial carcinoma of the uterus. *Cancer Res.* 51:5632-5.

#### Example 21: Ribosomal Protein S14 (RPS14) - Target Gene VARIA326

##### *Ribosomal protein S14 is essential for cell growth*

Human ribosomal protein S14 (RPS14) is one of ~80 unique protein constituents of the mammalian ribosome. Many of the protein subunits of ribosomes, the protein making machines of all cells, are highly conserved throughout prokaryotic and eukaryotic evolution (1). For example, human RPS14 protein is 100% amino acid identical to hamster S14 protein, 72% identical to yeast rp59 protein and 43% identical to *E. Coli* ribosomal protein S11 (2,3). Mammalian S14 and yeast rp59 are components of the 40S ribosomal subunit while *E. coli* S11 is part of the corresponding bacterial S30 subunit. Thus human RPS14 is a ribosomal component fixed early in evolution.

There are many antibiotics and eukaryotic cell poisons that act by inhibiting ribosome function (reviewed in ref. 4). One such drug is emetine, which inhibits protein translation by interacting with the eukaryotic RPS14 subunit to prevent elongation

240

232/116

factor dependent translocation of peptidyl-tRNAs bound to eukaryotic ribosomes in vitro (4).

Chinese hamster ovary (CHO) cell lines resistant to emetine have been shown to contain mutant RPS14 loci (also referred to as the EMTB locus) (5). Such lines have been used to investigate the effects of mutant RPS14 on ribosome function (5-8). Human-CHO cell hybrids are emetine-sensitive, indicating that the EMTB/RPS14 mutation is recessive in CHO cells. This is apparently because arrest of protein synthesis in half of ribosomes blocks translation of all polysomic mRNAs by blocking any functional ribosomes upstream of frozen mutant ribosomes. RPS14 appears to contribute to the structural integrity of the 40S subunit: 40S subunits containing mutant S14 protein are more easily dissociable in high ionic strength wash buffers (9). Ribosomal subunit genes are coordinately expressed in all cells and ribosomal proteins constitute a large fraction of the cell mass in all cell types.

*The human RPS14 gene has sequence variances*

Rhoads et al. reported the sequence of the human RPS14 gene and cDNA (3). The cDNA contains a 33 nucleotide 5' untranslated region, a 453 nt coding region and a 60 nt 3' untranslated region (including 12 nt of polyA). We undertook a systematic search for DNA sequence variance in the cDNA of RPS14 by analysing 36 unrelated individuals using the single strand conformation polymorphism technique. Primers were designed using the sequence of Rhoads et al. (GENBANK accession M13934, M13641; see ref. 3). SSCP analysis revealed 1 sequence variance, and subsequent DNA sequence analysis confirmed an A vs. G transition at nucleotide 183 of the coding sequence. (This change was noted as a difference between the cDNA and genomic sequences in ref. 3.)

As shown in the Target Summary Table, both alleles were detected in all major

populations surveyed, including North American Whites, North American Blacks, Hispanics, Chinese and Japanese.

*The human RPS14 gene maps to chromosome 5q23-q33*

5

Dana and Wasmuth (11) used Chinese hamster/human somatic cell hybrids to map the RPS14 gene (designated EMTB) to 5q23-5q35. Later Nakamichi et al. (12) placed the RPS14 gene on the segment 5q23-q33 using similar techniques.

10

*Chromosome band 5q23-q33 is a site of frequent loss of heterozygosity.* There have been many studies of LOH on 5q, particularly the 5q21-q22 region where the Adenomatous Polyposis Coli (APC) tumor suppressor gene lies. The most extensively studied cancers are those of the gastrointestinal tract, lung and ovary. The available data on the 5q23-q33 region just distal to APC (where RPS14 lies), suggests that LOH occurs in this region at a frequency of ~30% in cervical cancer (13), 20-40% in colon cancer (14,15), 30-50% in ovarian cancer (16,17), 38% in stomach cancer (18) and 23% in testicular cancer (19). There is also evidence for LOH in head and neck, lung, and liver cancers.

15

20

## References

25

1. Chambliss, G., Craven, G.R., Davies, J., et al., editors, Ribosomes: Structure, Function and Genetics, University Park Press, Baltimore, 1980.
2. Chen, I.-T., Dixit, A., Rhoads, D.D. and D.J. Roufa (1986) Homologous ribosomal proteins in bacteria, yeast and humans. *Proc. Natl. Acad. Sci. U.S.A.* 83: 6907-6911.
3. Rhoads, D. D.; Dixit, A.; Roufa, D. J. (1986) Primary structure of human ribosomal protein S14 and the gene that encodes it. *Molec. Cell. Biol.* 6: 2774-2783.
4. Vazquez, D. (1979) Molecular Biology and Biophysics, vol. 30, Inhibitors of Protein Synthesis. Springer-Verlag, Berlin.

5. Wasmuth, J.J. (1985) Chinese hamster cell protein synthesis mutants. In Gottesman, M., ed. *Molecular Cell Genetics*, pp. 397-421.
6. Rhoads, D.D. and D.J. Roufa (1985) Emetine resistance in Chinese hamster cells: structures of wild-type and mutant ribosomal protein AS14 mRNAs. *Mol. Cell Biol.* 5: 1655-1659.
7. Madjar, J.J., Nielsen-Smith, K., Frahm, M. and D. Roufa (1982) Emetine resistance in Chinese hamster ovary cells is associated with an altered ribosomal protein S14 mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 79: 1003-1007.
8. Dana, S. L., Chang, S. and J.J. Wasmuth (1985) Synthesis and incorporation of human ribosomal protein S14 into functional ribosomes in human-Chinese hamster cell hybrids containing human chromosome 5: human RPS14 gene is the structural gene for ribosomal protein S14. *Somat. Cell Molec. Genet.* 11: 625-631.
9. Madjar, J.-J., Frahm, M., McGill, S. and D.J. Roufa (1983) *Molec. Cell. Biol.* 3: 190-197.
10. Mount, S. (1982) A catalogue of splice junction sequences. *Nucleic Acids Research* 19: 459-472.
11. Dana, S. and J.J. Wasmuth (1982) Selective linkage disruption in human-Chinese hamster cell hybrids: deletion mapping of the leuS, hexB, emtB, and chr genes on human chromosome 5. *Molec. Cell. Biol.* 2: 1220-1228.
12. Nakamichi, N. N.; Kao, F.-T.; Wasmuth, J.; Roufa, D. J. (1986) Ribosomal protein gene sequences map to human chromosomes 5, 8 and 17. *Somat. Cell. Molec. Genet.* 12: 225-236.
13. Mitra AB, Murty VV, Li RG, Pratap M, Luthra UK, Chaganti RS. (1994) Allelotype analysis of cervical carcinoma. *Cancer Res.* 54:4481-7.
14. *Japanese Journal of Cancer Research* 82: 1003.
15. Cunningham C, Dunlop MG, Wyllie AH, Bird CC. (1993) Deletion mapping in colorectal cancer of a putative tumor suppressor gene in 8p22-p21.3. *Oncogene.* 8:1391-6.
16. *British Journal of Cancer* 69: 429.

17. Weitzel J.N., Patel J., Smith D.M., Goodman A., Safaii H., Ball H.G. (1994) Molecular genetic changes associated with ovarian cancer. *Gynecol. Oncol.* 55:245-52.
18. *Genes, Chromosomes and Cancer* 3: 468
19. Murty VV, Bosl GJ, Houldsworth J, et al. (1994) Allelic loss and somatic differentiation in human male germ cell tumors. *Oncogene*. 9:2245-51.

**Example 22: Eukaryotic Initiation Factor 5A (eIF-5A) - Target Gene VARIA351**

*Initiation Factor 5A is essential for cell growth*

Human Initiation Factor 5A (eIF-5A), formerly named Initiation Factor 4D, is an 18-kD protein which promotes formation of the first peptide bond in *in vitro* translation systems - hence the name 'initiation factor' (1,2); however, the full physiological role of eIF-5A is not understood. Inhibition of eIF 5A formation blocks proliferation in all tested cell types (3); the presence of functional eIF 5A has been shown to correlate with the onset of DNA replication (4) - perhaps due to eIF 5A dependent translation of mRNAs encoding proteins necessary for DNA replication (3), and eIF-5A is an essential co-factor for HIV-1 Rev protein (5).

eIF 5A is an unusual protein: one of its lysine residues (amino acid 50) is modified by transfer and hydroxylation of the butylamino-group from the polyamine spermidine to form hypusine, a post translational modification unique to eIF 5A. All of the biological activities of eIF 5A are abrogated in the absence of the hypusine modification, as demonstrated by pharmacological inhibition of hypusine formation in human cell lines (3) and by site directed mutagenesis of the modified lysine residue in the yeast enzyme (6). There are two enzymes responsible for hypusine formation, one of which, deoxyhypusyl hydroxylase, can be inhibited with the drug mimosine (3), providing a convenient pharmacological inhibitor of eIF 5A formation.

The genome of the yeast *Saccharomyces cerevisiae* encodes two eIF 5A genes. Disruption of one (form A) slows growth, disruption of the other (form B) arrests growth and strains with both forms disrupted are non-viable (6). The yeast A form substitutes for human eIF 5A in the mammalian methionyl-puromycin synthesis assay (6), while the human gene complements eIF 5A disrupted yeast (7). eIF 5A is a highly conserved protein, with counterparts in archaea, bacteria and eukaryotes. The yeast proteins are ~63% identical to the human protein (6).

*The human eIF 5A gene and mRNA have sequence variances*

Smit-McBride, et al. reported the sequence of a human cDNA encoding eIF-5A (8) and Koettwitz et al. (8) later reported the sequence of the active eIF 5A gene, which contains three introns (GenBank accession U17969). A composite sequence made from the cDNA and genomic versions is 1309 nucleotides long and contains a 5' untranslated region of 145 nucleotides, a 462 nt coding region and a 702 nt 3' untranslated region (see annotated sequence). We undertook a systematic search for DNA sequence variance in the cDNA of eIF 5A by analysing 36 unrelated individuals using the single strand conformation polymorphism technique. Primers were designed for amplification. SSCP analysis revealed 2 sequence variances, and subsequent DNA sequence analysis confirmed an A vs. G transition at nucleotide 623 and a T vs. C transition at nucleotide 1012, both in the 3' untranslated sequence.

Neither sequence variance affects the protein coding sequence, however nucleotide 623 is one nucleotide away from a splice acceptor site at position 622, and could therefore be targeted by an oligonucleotide intended to abrogate splicing in an allele specific manner. The second exonic nucleotide (+2 position) of a splice acceptor site is not highly conserved, nonetheless the A vs. G transition at nucleotide 623 may affect the mechanics of splicing.

245

232/116

As shown in the Target Summary Table, both alleles were detected in all major populations surveyed, including North American Whites, North American Blacks, Hispanics, Arabs, Indians and Japanese, except only the nucleotide 1012 variance was detected in the four Chinese surveyed. The overall frequency of heterozygotes was 37% for the nucleotide 623 sequence variance and 52% for the nucleotide 1012 sequence variance.

*The human eIF 5A gene maps to chromosome 17p13-p12*

Steinkasserer et al. (1995) mapped the eIF 5A gene to 17p13-p12 by fluorescence *in situ* hybridization (9). Three eIF 5A pseudogenes were mapped to 10q23, 17q25 and 19q13.

*Chromosome band 17p13-p12 is a site of frequent loss of heterozygosity.* There have been many studies of LOH on 17p, particularly the 17p13 region where the p53 tumor suppressor gene maps. Virtually all cancer types have been surveyed for LOH in this area, with particularly extensive studies of breast, colon, ovarian, and stomach cancers. These studies report LOH in approximately 40-60% of breast cancers (10-18), 50-70% of colon cancers (19-25), 25-75% of ovarian cancers (26-30), 20-60% of stomach cancers (31-34), 20-50% of brain cancers (35,36), 45-70% of esophageal cancers (37), 35-65% of non-small cell lung cancers (38,39) and 100% of small cell lung cancers, 15-50% of cervical cancers, 30-80% of head and neck cancers, 20-60% of liver cancers, over 50% of sarcomas and 10-30% of a variety of other cancer types.

**References**

1. Wolff, E.C., Park, M.H. and J.E. Folk (1990) *Journal of Biological Chemistry* 265: 4793-4799.
2. Park, M.H., Wolff, E.C. and J.E. Folk (1993) Hypusine: its post-translational formation in eukaryotic translation factor 5A and its potential role in cellular



regulation. *Biofactors* 4: 95-104.

3. Hanauske-Abel, H.M., Park, M.-H., Hanauske, A.-R., et al. (1994) Inhibition of the G1-S transition of the cell cycle by inhibitors of deoxyhypusine hydroxylation. *Biochimica et Biophysica Acta* 1221: 115-124.

5 4. Hanauske-Abel, H.M., Slowinska, B., Zagulska, S., et al. (1995) Detection of a subset of polysomal mRNAs associated with modulation of hypusine formation at the G1-S boundary. Proposal of a role for eIF 5A in onset of DNA replication. *FEBS Lett.* 366: 92-98.

10 5. Ruhl, M., Himmelsbach, M., Bahr, G.M., et al. (1993) Eukaryotic initiation factor 5A is a cellular target of the HIV-1 Rev activation domain mediating trans-activation. *J. Cell Biol.* 123:1309-1320.

6. Schnier, J., Schwelberger, H.G., Smit-McBride, Z., et al. (1991) Translation initiation factor 5A and its hypusine modification are essential for cell viability in the yeast *Saccharomyces Cerevisiae*. *Molecular and Cellular Biology* 11: 3105-3114.

15 7. Koettnitz, K., Wohl, T., Kappel, B., Lottspeich, F., Hauber, J. and D. Bevec (1995) Identification of a new member of the human eIF-5A gene family. *Gene* 159: 283-284.

20 8. Smit-McBride, Z., Dever, T.E., Hershey, J.W.B., et al. (1989) Sequence determination and cDNA cloning of eukaryotic initiation factor 4D, the hypusine containing protein. *Journal of Biological Chemistry* 264: 1578-1583.

9. Steinkasserer, A.; Jones, T.; Sheer, D.; Koettnitz, K.; Hauber, J. and D. Bevec (1995) The eukaryotic cofactor for the human immunodeficiency virus type 1 (HIV-1) rev protein, eIF-5A, maps to chromosome 17p12-p13: three eIF-5A pseudogenes map to 10q23.3, 17q25, and 19q13.2. *Genomics* 25: 749-752.

25 10. Comelis RS, van Vliet M, Vos CB, et al. (1994) Evidence for a gene on 17p13.3, distal to TP53, as a target for allele loss in breast tumors without p53 mutations. *Cancer Res.* 54:4200-6.

11. Lindblom A, Skoog L, Rotstein S, Werelius B, Larsson C, Nordenskjold M. (1993) Loss of heterozygosity in familial breast carcinomas. *Cancer Res.* 53:4356-61.

12. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* 51:5794-9.
13. Singh S, Simon M, Meybohm I, et al. (1993) Human breast cancer: frequent p53 allele loss and protein overexpression. *Hum Genet.* 90:635-40.
- 5 14. Thorlacius S, Borresen AL, et al. (1993) Somatic p53 mutations in human breast carcinomas in an Icelandic population: a prognostic factor. *Cancer Res.* 53:1637-41.
15. Tsuda H, Hirohashi S. (1994) Association among p53 gene mutation, nuclear accumulation of the p53 protein and aggressive phenotypes in breast cancer. *Int J Cancer.* 57:498-503.
- 10 16. Watatani M, Nagayama K, Imanishi Y, et al. (1993) Genetic alterations on chromosome 17 in human breast cancer: relationships to clinical features and DNA ploidy. *Breast Cancer Res Treat.* 28:231-9.
17. Chen LC, Neubauer A, Kurisu W, et al. (1991) Loss of heterozygosity on the short arm of chromosome 17 is associated with high proliferative capacity and DNA aneuploidy in primary human breast cancer. *Proc Natl Acad Sci U S A.* 88:3847-51.
- 15 18. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* 51:5794-9.
19. Burner GC, Rabinovitch PS, Haggitt RC, et al. (1992) Neoplastic progression in ulcerative colitis: histology, DNA content, and loss of a p53 allele [see comments].
- 20 *Gastroenterology.* 103:1602-10.
20. Cunningham C, Dunlop MG, Wyllie AH, Bird CC. (1993) Deletion mapping in colorectal cancer of a putative tumor suppressor gene in 8p22-p21. *Oncogene.* 8:1391-6
- 25 21. Kikuchi-Yanoshita R, Konishi M, Ito S, et al. (1992) Genetic changes of both p53 alleles associated with the conversion from colorectal adenoma to early carcinoma in familial adenomatous polyposis and non-familial adenomatous polyposis patients. *Cancer Res.* 52:3965-71.
22. Yin J, Harpaz N, Tong Y, et al. (1993) p53 point mutations in dysplastic and cancerous ulcerative colitis lesions. *Gastroenterology.* 104:1633-9.

23. Iacopetta B, DiGrandi S, Dix B, et al. (1994) Loss of heterozygosity of tumor suppressor gene loci in human colorectal carcinoma. *Eur J Cancer*. 5:664-70.
24. Law DJ, Olschwang S, Monpezat JP, et al. (1988) Concerted nonsyntenic allelic loss in human colorectal carcinoma. *Science*. 241:961-5.
- 5 25. Lothe RA, Nakamura Y, Woodward S, Gedde DT, Jr., White R. (1988) VNTR (variable number of tandem repeats) markers show loss of chromosome 17p sequences in human colorectal carcinomas. *Cytogenet Cell Genet*. 48:167-9.
26. Foulkes WD, Stamp GW, Afzal S, et al. (1995) MDM2 overexpression is rare in ovarian carcinoma irrespective of TP53 mutation status. *Br J Cancer*. 72:883-8.
- 10 27. Phillips NJ, Ziegler MR, Radford DM, et al. (1996) Allelic deletion on chromosome 17p13.3 in early ovarian cancer. *Cancer Res*. 56:606-11.
28. Foulkes WD, Black DM, Stamp GW, Solomon E, Trowsdale J. (1993) Very frequent loss of heterozygosity throughout chromosome 17 in sporadic ovarian carcinoma. *Int J Cancer*. 54:220-5.
- 15 29. Gallion HH, Powell DE, Morrow JK, et al. (1992) Molecular genetic changes in human epithelial ovarian malignancies [see comments]. *Gynecol Oncol*. 47:137-42.
30. Phillips N, Ziegler M, Saha B, Xynos F. (1993) Allelic loss on chromosome 17 in human ovarian cancer. *Int J Cancer*. 54:85-91.
31. Seruca R, David L, Castedo S, Veiga I, Borresen AL, Sobrinho-Simoes M. (1994) p53 alterations in gastric carcinoma: a study of 56 primary tumors and 204 nodal metastases. *Cancer Genet Cytogenet*. 75:45-50.
- 20 32. Kim CJ, Kim WH, Kim CW, Lee JB, Lee CK, Kim YL. (1995) Detection of 17p loss in gastric carcinoma using polymerase chain reaction. *Lab Invest*. 72:232-6.
33. Ranzani GN, Renault B, Pellegata NS, et al. (1993) Loss of heterozygosity and K-ras gene mutations in gastric cancer. *Hum Genet*. 92:244-9.
- 25 34. Sano T, Tsujino T, Yoshida K, et al. (1991) Frequent loss of heterozygosity on chromosomes 1q, 5q, and 17p in human gastric carcinomas. *Cancer Res*. 51:2926-31.
35. Frankel RH, Bayona W, Koslow M, Newcomb EW. (1992) p53 mutations in human malignant gliomas: comparison of loss of heterozygosity with mutation

frequency. *Cancer Res.* 52:1427-33.

36. Hermanson M, Funa K, Koopmann J, et al. (1996) Association of loss of heterozygosity on chromosome 17p with high platelet-derived growth factor alpha receptor expression in human malignant gliomas. *Cancer Res.* 56:164-71.

5 37. Aoki T, Mori T, Du X, Nisihira T, Matsubara T, Nakamura Y. (1994) Allelotype study of esophageal carcinoma. *Genes Chromosomes Cancer.* 10:177-82.

38. Tsuchiya E, Nakamura Y, Weng SY, et al. (1992) Allelotype of non-small cell lung carcinoma--comparison between loss of heterozygosity in squamous cell carcinoma and adenocarcinoma. *Cancer Res.* 52:2478-81.

10 39. Hiyama K, Ishioka S, Shirotani Y, et al. (1995) Alterations in telomeric repeat length in lung cancer are associated with loss of heterozygosity in p53 and Rb. *Oncogene.* 10:937-44.

15 **Example 23: Replication Protein A, 32 kDa Subunit (RPA32) - Target Gene VARIA402**

*The human RPA32 gene encodes a protein essential for cell survival*

20 Replication Protein A (RPA; also known as Replication Factor A, Activator 1, Single Strand Binding Protein or SSB) is a heterotrimeric protein which participates in DNA replication, homologous recombination and nucleotide excision repair (1-3). The evidence that RPA is an essential protein comes from *in vitro* and *in vivo* data.

25 DNA replication is essential for cell proliferation, as discussed above for RPA70.

The best studied function of RPA32 is in DNA replication. Because of the complexity of DNA replication in higher eukaryotic genomes, the small genome of the papovavirus SV40 has been used as a model system to study DNA replication in human cell extracts. In the 1980s several research groups

250

232/116

developed cell free systems to study DNA replication using SV40 chromosomes as templates (4-8). An effort to identify the minimal set of factors required for DNA replication led to the discovery of RPA. Subsequent work proved that each of the three subunits of RPA is essential for DNA replication (9,10). This was proved in several ways, including by using antibodies to various constituents of the replication complex. Anti-RPA32 antibodies inhibit DNA replication, providing clear *in vitro* evidence for the essential function of this subunit of RPA in human DNA replication (10).

The yeast *S. cerevisiae* has a trimeric replication protein A which is structurally and functionally homologous to the human protein. It consists of three subunits similar in size to the human subunits. All three yeast subunits have been disrupted and each disruption produces non-viable yeast (9).

*The human RPA32 gene and mRNA are polymorphic.*

The published cDNA for the 32 kD subunit of Replication Protein A is 1512 nucleotides long and includes a 5' untranslated segment of 77 nucleotides, followed by a protein coding region of 810 nucleotides and a 3' untranslated region of 625 nucleotides (10). We undertook a systematic search for DNA polymorphism by analysing the RPA32 cDNA from 36 unrelated individuals using the single strand conformation polymorphism technique (described in the methods section). Primers were designed using the sequence of Erdile et al. (GenBank accession J05249; see ref. 10). SSCP analysis revealed 2 variances, one of which was sequenced. Sequencing revealed a G vs. A transition at nucleotide 40 of the 5' untranslated region. Four of 36 individuals were heterozygotes, all of them Caucasians. Thus the allele frequency is 25% (4/16) in North American Whites, while no heterozygosity was detected in other populations (see Target Summary sheet).

*The RPA32 gene maps to chromosome 1p35*

The gene for RPA32 was mapped to chromosome band 1p35 by *in situ* hybridization, somatic cell hybrid analysis and yeast artificial chromosome mapping (11,12). Only one locus was detected by all methods.

5

*Chromosome band 1p35 is a site of frequent loss of heterozygosity.* The short arm of chromosome 1 is comparatively well investigated for allele loss, especially in breast and colon cancers. Studies of the 1p35 region show LOH in 15-40% of breast cancers (13,14), ~50% of gliomas (a brain cancer subtype) (15), 20-70% of colon cancers (16,17), ~50% of stomach cancers (18), ~20% of lung cancers (19) and 10-30% of ovarian cancers. High frequency LOH has been detected in several uncommon cancers such as pheochromocytoma (50-80%) and neuroblastoma (~50%).

10

## References

15

1. Erdile, L. F., et al. Characterization of a cDNA encoding the 70-kDa single-stranded DNA-binding subunit of human replication protein A and the role of the protein in DNA replication. [published erratum appears in *J. Biol. Chem.* 1993 Jan 25;268(3):2268]. *J. Biol. Chem.* 266.18 (1991): 12090-8.

20

2. Jones, K. A., et al. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* 48.1 (1987): 79-89.

3. He, Z., et al. RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* 374.6522 (1995): 566-9.

25

4. Challberg, M. D., and T. J. Kelly. Eukaryotic DNA replication: viral and plasmid model systems. *Annu Rev Biochem* 51 (1982): 901-34.

5. Wold, M. S., et al. Identification of cellular proteins required for simian virus 40 DNA replication. *Journal Biological Chemistry* 264.5 (1989): 2801-9.

6. Kelly, T. J. DNA replication in mammalian cells: insights from the SV40 model system. *Harvey Lecture* 85 (1989): 173-88.

7. Hurwitz, J., Dean, F.B., Kwong, A.D and S.-H. Lee (1990) *Journal of Biological Chemistry* 265: 18043-18046.

8. Stillman, B. (1992) Initiation of chromosome replication in eukaryotic cells. *Harvey Lecture* 88: 115-40.

5 9. Brill, S.J. and B. Stillman (1991) Replication factor-A from *Saccharomyces cerevisiae* is encoded by three essential genes coordinately expressed at S phase. *Genes and Development* 5: 589-1600.

10. Erdile, L. F., M. S. Wold, and T. J. Kelly. The primary structure of the 32-kDa subunit of human replication protein A. *J Biol Chem* 265.6 (1990): 3177-82.

10 11. Ozawa, K., Dean, F., et al. (1993) Mapping of the 70 kDa 34kDa and 11kDa subunit genes of the human multimeric single-stranded DNA binding protein (hSSB/RPA) to chromosome bands 17p13, 1p35-p36.1 and 7p21-p22. *Cell Struct Funct* 18: 221-230.

15 12. Umbricht, C. B., et al. High-resolution genomic mapping of the three human replication protein A genes (RPA1, RPA2, and RPA3). *Genomics* 20.2 (1994): 249-57.

13. Bieche I, Champeme MH, Matifas F, Cropp CS, et al. (1993) Two distinct regions involved in 1p deletion in human primary breast cancer. *Cancer Res.* 53:1990-4.

20 14. Borg A, Zhang QX, Olsson H, Wenngren E. (1992) Chromosome 1 alterations in breast cancer: allelic loss on 1p and 1q is related to lymphogenic metastases and poor prognosis. *Genes Chromosomes & Cancer.* 5:311-20.

15. Reifenberger, J., Reifenberger, G., Liu, L., et al. (1994) Molecular genetic analysis of oligodendroglial tumors shows preferential allelic deletions on 19q and 1p. *American Journal of Pathology* 145: 1175-1190.

25 16. (1990) *Cancer Research* 50: 7232.

17. Lothe RA, Nakamura Y, Woodward S, Gedde DT, Jr., White R. (1988) VNTR (variable number of tandem repeats) markers show loss of chromosome 17p sequences in human colorectal carcinomas. *Cytogenet Cell Genet.* 48:167-9.

18. Ezaki, T., Yanagisawa, A., Ohta, K., et al. (1996) Deletion mapping on

chromosome 1p in well-differentiated gastric cancer. *British Journal of Cancer* 73: 424-428.

19. Hiyama K, Ishioka S, et al. (1995) Alterations in telomeric repeat length in lung cancer are associated with loss of heterozygosity in p53 and Rb. *Oncogene*. 10:937-44.

5

**Example 24: Replication Protein A, 70 kD subunit (RPA70) - Target Gene VARIA401**

10

*The human RPA70 gene encodes a protein essential for cell survival*

Replication Protein A (also known as Replication Factor A, Activator or Single Strand Binding protein [SSB]) is a heterotrimeric protein which participates in DNA replication, homologous recombination and nucleotide excision repair (1-3). The evidence that RPA is an essential protein comes from *in vitro*, *in vivo* and evolutionary data.

15

20

DNA replication is essential for cell proliferation, and a variety of antiproliferative drugs act, at least in part, by inhibiting DNA replication. Such drugs include nucleotide analogs that block DNA polymerases, such as 2',3' dideoxy NTPs and 3' deoxy ATP (cordycepin); inhibitors that bind to or modify DNA such as intercalating agents, DNA crosslinking drugs or alkylating agents, and inhibitors that bind to polymerases and replication proteins such as topoisomerase inhibitors like the epipodophyllotoxins, which prevent DNA unwinding necessary for replication (and transcription) and antibiotics which bind to polymerases such as arylhydrazino-pyrimidines.

25

The best studied function of RPA70 is in DNA replication. Because of the complexity of DNA replication in higher eukaryotic genomes, the small genome of the papovavirus SV40 has been used as a model system to study DNA replication in human cell extracts. In the 1980s several research groups



254

232/116

developed cell free systems to study DNA replication using SV40 chromosomes as templates (4-8). These studies, in seeking to identify the minimal set of factors required for DNA replication, led to the discovery of replication protein A. Subsequent work proved that each of the three subunits of RPA is essential for DNA replications. This was proved in several ways, including by using antibodies to various constituents of the replication complex. These antibodies are effectively inhibitors of RPA70. Anti-RPA70 antibody mediated abrogation of DNA replication provides clear *in vitro* evidence for the essential function of RPA70 in human DNA replication (10). The yeast *S. cerevisiae* has a trimeric replication protein A which is structurally and functionally homologous to the human protein. It consists of three subunits similar in size to the human subunits. The yeast 70 kDa subunit is 31% identical and 75% similar (including conserved amino acids) to its human counterpart (1). All three yeast subunits have been disrupted and each disruption produces non-viable yeast. The yeast 70 kD protein is also a single stranded DNA binding protein.

Single stranded DNA binding proteins (SSBs) are required for DNA replication in a wide variety of organisms, including bacteriophage, bacteria and some DNA viruses of higher eukaryotes. Recently the crystal structure of the DNA binding domain of human RPA was solved and found to be remarkably similar in three dimensional shape to the bacteriophage single stranded DNA binding proteins Pf3 and gene V from  $\phi$ 1 phage.

*The human RPA70 gene, mRNA and protein have sequence variances*

The published cDNA for the 70 kD subunit of Replication Protein A is 2393 nucleotides long and includes a 5' untranslated segment of 69 nucleotides, followed by a protein coding region of 1848 nucleotides and a 3' untranslated region of 476 nucleotides (1). We undertook a systematic search for DNA polymorphism by

255

232/116

analyzing the RPA70 cDNA from 36 unrelated individuals using the single strand conformation polymorphism technique (described in the methods section). Primers were designed using the sequence of Erdile et al. (GenBank accession M63488; see ref. 1). SSCP analysis revealed 5 variances, and subsequent DNA sequence analysis of those variances led to identification of four additional variances. SSCP revealed the variances at nucleotides 81 (G vs. A), 1120 (A vs. G), 1674 (T vs. C), 2050 (T vs. C) and 2297, where an insertion/deletion variance of one C nucleotide was observed (8 vs. 9 C's in a row). In the course of sequencing around the nucleotide 2297 polymorphism an additional variance was detected at nucleotide 2341 (A vs. G). Also, while sequencing additional Swedish individuals around nucleotide 1120 two new variances were observed at nucleotides 1124 and 125 (both C vs. T). Finally, in three individuals sequenced for the 2050 variance we noted a difference from the published sequence at nucleotide 2046: we detect 3 T's while the published clone shows just two. This difference may represent another insertion/deletion polymorphism. Five of the nine detected variances are in the coding sequence while four are in the 3' untranslated region.

The frequency of heterozygotes for the five SSCP positive variances ranged from 25-42% among the 36 individuals tested. The small number of individuals genotyped for the other four variances precludes definitive assessment of heterozygosity rates. Some of the polymorphisms appear to occur more commonly in certain racial or ethnic groups (see Target Summary sheet for details). For example, only one of the variances (nt 1674) was detected in Japanese individuals. In general, higher levels of polymorphism were detected in North American Whites than in other groups. The nucleotide 1120 polymorphism, for instance, was heterozygous in 9/36 individuals overall (25%), but in 8/16 North American Whites (50%).

The RPA70 cDNA encodes a 616 amino acid protein. The nucleotide 1120 and 1124 variances result in amino acid substitutions at residues 351 and 352, the former an alanine-threonine exchange (approximately 50% of caucasians are heterozygotes) and

256

232/116

the latter a serine-phenylalanine exchange (rare in the populations tested). In the recently published crystal structure of the DNA binding segment of RPA70 (amino acids 181-422) it is possible to place residue 351 in the second of two tandemly arrayed DNA binding domains (domain B; see ref. 10). Domain B extends from residue I305 to N402, thus the variant residue 351 is in the middle. The published structure is a co-crystal of RPA70 amino acids 181-422 complexed to octadeoxycytosine. Several RPA70 residues contact the oligonucleotide (Figure 4 of ref. 11), including amino acids K343 and T359, which lie 8 residues away from the polymorphism in either direction. Modeling the two variant forms of the protein using the atomic coordinates deposited in the Protein Data Bank (1JMC) should clarify the structural consequences of the alanine-threonine variance. Residue 351 lies in the center of a 50 amino acid segment of the protein that is relatively poorly conserved between yeast and man: 11 of the 50 residues are identical and 25 more are conservative substitutions. Towards the C terminus there is strong conservation: starting 25 residues C-terminal of the polymorphism, 27 of the next 37 residues are identical between yeast and man. Towards the N terminus there is ~30% conservation. Both yeast and human 70 kD RPA subunits contain putative C4-type zinc finger motifs at positions ~480-500.

*The RPA70 gene maps to chromosome 17p13.3*

The gene for RPA70 has been mapped to chromosome band 17p13.3 by *in situ* hybridization (12). Only one locus was detected.

*Chromosome band 17p13.3 is a site of frequent loss of heterozygosity.* RPA70 lies just telomeric to the TP53 tumor suppressor gene which is located in cytogenetic band 17p13.1. This region of chromosome 17 is extremely well investigated for allele loss. In general, studies report LOH in approximately 40-60% of breast cancers (13-21), 50-70% of colon cancers (22-28), 25-75% of ovarian cancers (29-33), 20-60% of stomach cancers (34-37), 20-50% of brain cancers (38,39), 45-70% of esophageal cancers (40),

35-65% of non-small cell lung cancers (41,42) and 100% of small cell lung cancers, 15-50% of cervical cancers, 30-80% of head and neck cancers, 20-60% of liver cancers, over 50% of sarcomas and 10-30% of a variety of other cancer types.

#### 5 *Assays developed for RPA: Protein and DNA contacts*

Human cDNAs encoding all 3 subunits (70, 34 and 11 kD) of RPA have been cloned and expressed in *E. coli* and in insect cells via baculovirus vectors. The bacterially expressed 70 kDa protein is indistinguishable from its purified human counterpart immunologically and in several functional assays (see Table below). There is good evidence that the 70 kD subunit of RPA interacts with a number of different molecules. A partial list would include the 34 and 11 kD subunits of RPA, DNA, the xeroderma pigmentosum damage recognition and endonuclease proteins XPA and XPG, and DNA polymerase  $\alpha$ -primase. These experimentally proven contacts (and almost certainly others) may constrain the topology of the protein in ways that have implications for inhibitor design. In summary a broad array of assays exists to screen for small molecule inhibitors of RPA (possibly including modified nucleotides), that act via competitive, allosteric or protein-protein blocking mechanisms.

**Table 4**

#### 20 **Assays and reagents available for RPA inhibitor screening**

ASSAY	RPA 70 kD, Assay Systems	
	Purified Human Protein	Purified Bacterial or Baculovirus Protein
Immunoreactivity	X	X
Single stranded DNA binding	X	X
DNA Polymerase $\alpha$ primase	X	X

258

232/116

DNA strand exchange	X	X
Nucleotide excision repair	X	X
Support SV40 Replication	X	X

## 5      References

1. Erdile, L. F., et al. Characterization of a cDNA encoding the 70-kDa single-stranded DNA-binding subunit of human replication protein A and the role of the protein in DNA replication. [published erratum appears in *J. Biol. Chem.* 1993 Jan 25;268(3):2268]. *J. Biol. Chem.* 266.18 (1991): 12090-8.
2. Jones, K. A., et al. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* 48.1 (1987): 79-89.
3. He, Z., et al. RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* 374.6522 (1995): 566-9.
4. Challberg, M. D., and T. J. Kelly. Eukaryotic DNA replication: viral and plasmid model systems. *Annu Rev Biochem* 51 (1982): 901-34.
5. Wold, M. S., et al. Identification of cellular proteins required for simian virus 40 DNA replication. *Journal Biological Chemistry* 264.5 (1989): 2801-9.
6. Kelly, T. J. DNA replication in mammalian cells: insights from the SV40 model system. *Harvey Lecture* 85 (1989): 173-88.
7. Hurwitz, J., Dean, F.B., Kwong, A.D and S.-H. Lee (1990) *Journal of Biological Chemistry* 265: 18043-18046.
8. Stillman, B. (1992) Initiation of chromosome replication in eukaryotic cells. *Harvey Lecture* 88: 115-40.
9. Heyer, W. D., et al. An essential *Saccharomyces cerevisiae* single-stranded DNA binding protein is homologous to the large subunit of human RP-A. *EMBO Journal* 9.7 (1990): 2321-9.
10. Erdile, L. F., M. S. Wold, and T. J. Kelly. The primary structure of the 32-kDa subunit of human replication protein A. *J Biol Chem* 265.6 (1990): 3177-82.
11. Bochkarev, A., Pfuetzner, R.A., Edwards, A.M. and L. Frappier (1997) Structure

of the single stranded DNA binding domain of replication protein A bound to DNA. *Nature* 385: 176-181.

12. Umbricht, C. B., et al. High-resolution genomic mapping of the three human replication protein A genes (RPA1, RPA2, and RPA3). *Genomics* 20.2 (1994): 249-57.

13. Cornelis RS, van Vliet M, Vos CB, et al. (1994) Evidence for a gene on 17p13.3, distal to TP53, as a target for allele loss in breast tumors without p53 mutations. *Cancer Res.* 54:4200-6.

14. Lindblom A, Skoog L, Rotstein S, Werelius B, Larsson C, Nordenskjold M. (1993) Loss of heterozygosity in familial breast carcinomas. *Cancer Res.* 53:4356-61.

15. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* 51:5794-9.

16. Singh S, Simon M, Meybohm I, et al. (1993) Human breast cancer: frequent p53 allele loss and protein over expression. *Hum Genet.* 90:635-40.

17. Thorlacius S, Borresen AL, et al. (1993) Somatic p53 mutations in human breast carcinomas in an Icelandic population: a prognostic factor. *Cancer Res.* 53:1637-41.

18. Tsuda H, Hirohashi S. (1994) Association among p53 gene mutation, nuclear accumulation of the p53 protein and aggressive phenotypes in breast cancer. *Int J Cancer.* 57:498-503.

19. Watatani M, Nagayama K, Imanishi Y, et al. (1993) Genetic alterations on chromosome 17 in human breast cancer: relationships to clinical features and DNA ploidy. *Breast Cancer Res Treat.* 28:231-9.

20. Chen LC, Neubauer A, Kurisu W, et al. (1991) Loss of heterozygosity on the short arm of chromosome 17 is associated with high proliferative capacity and DNA aneuploidy in primary human breast cancer. *Proc Natl Acad Sci U S A.* 88:3847-51.

21. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* 51:5794-9.

22. Burner GC, Rabinovitch PS, Haggitt RC, et al. (1992) Neoplastic progression in ulcerative colitis: histology, DNA content, and loss of a p53 allele [see comments].

*Gastroenterology*. 103:1602-10.

23. Cunningham C, Dunlop MG, Wyllie AH, Bird CC. (1993) Deletion mapping in colorectal cancer of a putative tumor suppressor gene in 8p22-p21.3. *Oncogene*. 8:1391-6

5 24. Kikuchi-Yanoshita R, Konishi M, Ito S, et al. (1992) Genetic changes of both p53 alleles associated with the conversion from colorectal adenoma to early carcinoma in familial adenomatous polyposis and non-familial adenomatous polyposis patients. *Cancer Res*. 52:3965-71.

10 25. Yin J, Harpaz N, Tong Y, et al. (1993) p53 point mutations in dysplastic and cancerous ulcerative colitis lesions. *Gastroenterology*. 104:1633-9.

26. Iacopetta B, DiGrandi S, Dix B, et al. (1994) Loss of heterozygosity of tumour suppressor gene loci in human colorectal carcinoma. *Eur J Cancer*. 5:664-70.

27. Law DJ, Olschwang S, Monpezat JP, et al. (1988) Concerted nonsyntenic allelic loss in human colorectal carcinoma. *Science*. 241:961-5.

15 28. Lothe RA, Nakamura Y, Woodward S, Gedde DT, Jr., White R. (1988) VNTR (variable number of tandem repeats) markers show loss of chromosome 17p sequences in human colorectal carcinomas. *Cytogenet Cell Genet*. 48:167-9.

29. Foulkes WD, Stamp GW, Afzal S, et al. (1995) MDM2 over expression is rare in ovarian carcinoma irrespective of TP53 mutation status. *Br J Cancer*. 72:883-8.

20 30. Phillips NJ, Ziegler MR, Radford DM, et al. (1996) Allelic deletion on chromosome 17p13.3 in early ovarian cancer. *Cancer Res*. 56:606-11.

31. Foulkes WD, Black DM, Stamp GW, Solomon E, Trowsdale J. (1993) Very frequent loss of heterozygosity throughout chromosome 17 in sporadic ovarian carcinoma. *Int J Cancer*. 54:220-5.

25 32. Gallion HH, Powell DE, Morrow JK, et al. (1992) Molecular genetic changes in human epithelial ovarian malignancies [see comments]. *Gynecol Oncol*. 47:137-42.

33. Phillips N, Ziegler M, Saha B, Xynos F. (1993) Allelic loss on chromosome 17 in human ovarian cancer. *Int J Cancer*. 54:85-91.

34. Seruca R, David L, Castedo S, Veiga I, Borresen AL, Sobrinho-Simoes M. (1994)

p53 alterations in gastric carcinoma: a study of 56 primary tumors and 204 nodal metastases. *Cancer Genet Cytogenet.* 75:45-50.

35. Kim CJ, Kim WH, Kim CW, Lee JB, Lee CK, Kim YL. (1995) Detection of 17p loss in gastric carcinoma using polymerase chain reaction. *Lab Invest.* 72:232-6.

5 36. Ranzani GN, Renault B, Pellegata NS, et al. (1993) Loss of heterozygosity and K-ras gene mutations in gastric cancer. *Hum Genet.* 92:244-9.

37. Sano T, Tsujino T, Yoshida K, et al. (1991) Frequent loss of heterozygosity on chromosomes 1q, 5q, and 17p in human gastric carcinomas. *Cancer Res.* 51:2926-31.

10 38. Frankel RH, Bayona W, Koslow M, Newcomb EW. (1992) p53 mutations in human malignant gliomas: comparison of loss of heterozygosity with mutation frequency. *Cancer Res.* 52:1427-33.

39. Hermanson M, Funa K, Koopmann J, et al. (1996) Association of loss of heterozygosity on chromosome 17p with high platelet-derived growth factor alpha receptor expression in human malignant gliomas. *Cancer Res.* 56:164-71.

15 40. Aoki T, Mori T, Du X, Nisihira T, Matsubara T, Nakamura Y. (1994) Allelotype study of esophageal carcinoma. *Genes Chromosomes Cancer.* 10:177-82.

41. Tsuchiya E, Nakamura Y, Weng SY, et al. (1992) Allelotype of non-small cell lung carcinoma--comparison between loss of heterozygosity in squamous cell carcinoma and adenocarcinoma. *Cancer Res.* 52:2478-81.

20 42. Hiyama K, Ishioka S, et al. (1995) Alterations in telomeric repeat length in lung cancer are associated with loss of heterozygosity in p53 and Rb. *Oncogene.* 10:937-44.

25 **Example 25: RNA Polymerase II, 220-kD subunit (RPOL2A) - Target Gene VARIA500**

*The human RPOL2A gene encodes a protein essential for cell survival*

DNA-dependent RNA polymerase II (also known as RPB1 or POLR2A), a complex



multisubunit enzyme, is responsible for the transcription of mRNA from all protein coding genes.

RNA polymerases are found in all cellular organisms. The subunit structure of RNA polymerases is highly conserved in eukaryotes. RNA polymerase acts in concert with as many as 50 other proteins in gene transcription (reviewed in ref. 1). See refs. 2 and 3 for a review of basal transcription by RNA polymerase II and recent progress in identifying and purifying transcription factors and cloning the genes that encode them.

Several subunits of *S. cerevisiae* RPOL2A have been disrupted, always resulting in non-viable yeast.

A variety of inhibitors of RNA polymerase are cytotoxic drugs, such as actinomycin D, which intercalates into double stranded DNA and blocks the movement of RNA polymerase; rifampicin binds the  $\beta$  subunit of *E. coli* RNA polymerase and blocks initiation of transcription. The best studied specific inhibitor of eukaryotic RPOL2A, however, is the potent mushroom toxin - amanitin, a cyclic octapeptide which binds with high affinity ( $K_d \sim 10^{-9}$  M) to RPOL2A. Several mutations conferring resistance to  $\alpha$ -amanitin have been characterized and they all map to the RPOL2A protein coding sequence. Recently  $\alpha$ -amanitin binding has been shown to trigger specific degradation of RPOL2A (4).

Damage to actively transcribed DNA is preferentially repaired by the transcription-coupled repair (TCR) system. TCR requires RNA pol II, but the mechanism by which repair enzymes preferentially recognize and repair DNA lesions on PolB II-transcribed genes is incompletely understood.

*The human RPOL2A gene and mRNA have sequence variances*

Wintzerith et al. and later Mita et al. cloned and sequenced the complete human gene

263

232/116

for RPOL2A (5, 6); the deduced amino acid sequences are identical. The RPOL2A gene contains 29 exons and spans about 32 kb of DNA. The cDNA sequence we evaluated is 6732 nucleotides long (see Annotated RPOL2A Sequence) and contains a 5' untranslated region of 386 nucleotides, a 5910 nucleotide coding region specifying 1970 amino acids, and a 436 nucleotide 3' untranslated region (see annotated sequence). We undertook a systematic search for DNA sequence variance in the cDNA of RPOL2A by analyzing 36 unrelated individuals using the single strand conformation polymorphism technique. Primers were designed for amplification. SSCP analysis revealed 10 sequence variances, and subsequent DNA sequence analysis confirmed a G vs. A transition at nucleotide 857, a C vs. T transition at nucleotide 1260, a C vs. T transition at nucleotide 1346, a C vs. T transition at nucleotide 1544, a C vs. T transition at nucleotide 1847, a C vs. T transition at nucleotide 2678, a C vs. T transition at nucleotide 3059, a C vs. T transition at nucleotide 3827, a T vs. C transition at nucleotide 6466 and a T vs. C transition at nucleotide 6557. The former seven sequence variances are in coding sequence and the latter two are in the 3' untranslated sequence. Only one of the ten sequence variances alters the protein coding sequence: the nucleotide 1260 alleles encode arginine (common) or cysteine (rare) at amino acid 292. Only 2/36 individuals surveyed are heterozygotes (6%), however both are North American Whites (2/16 = 12.5%) so further investigation of this population is required. The prevalence of heterozygotes for the other sequence variances varies from 3% to 50%, with 6 sequence variances above 22% (see RPOL2A Target Summary Sheet). The 6 common sequence variances are widely prevalent among all or nearly all the tested populations.

*The human RPOL2A gene maps to chromosome 17p13.105*

The human RPOL2A gene was initially assigned to the distal portion of the short arm of chromosome 17 (17pter-p12) by *in situ* hybridization and Southern analysis of DNA from human/rodent somatic cell hybrids (7, 8). Subsequent somatic cell hybrid studies narrowed the assignment to 17p13.105-p12 [vanTuinen and Ledbetter (1987)], which

was later confirmed by *in situ* hybridization to 17p13 (9).

*Chromosome band 17p13.1 is a site of frequent loss of heterozygosity* There have been many studies of LOH on 17p, particularly the 17p13.1 region where the p53 tumor suppressor gene maps. Virtually all cancer types have been surveyed for LOH in this area, with particularly extensive studies of breast, colon, ovarian, and stomach cancers. These studies report LOH in approximately 40-60% of breast cancers (10-18), 50-70% of colon cancers (19-25), 25-75% of ovarian cancers (26-30), 20-60% of stomach cancers (31-34), 20-50% of brain cancers (35,36), 45-70% of esophageal cancers (37), 35-65% of non-small cell lung cancers (38,39) and 100% of small cell lung cancers, 15-50% of cervical cancers, 30-80% of head and neck cancers, 20-60% of liver cancers, over 50% of sarcomas and 10-30% of a variety of other cancer types.

## References

1. Acker, J.; Mattei, M.-G.; Wintzerith, M.; Roeckel, N.; Depetris, D.; Vigneron, M.; Keding, C. (1994) Chromosomal localization of human RNA polymerase II subunit genes. *Genomics* 20: 496-499.
4. Buratowski, S. (1994) The basics of basal transcription by RNA polymerase II. *Cell* 77:1-3.
5. Cannizzaro, L. A., Emanuel, B. S., Cho, K. W. Y. and R. Weinmann (1986) The gene encoding the large subunit of human RNA polymerase II is located on the short arm of chromosome 17. *Am. J. Hum. Genet.* 38: 812-818.
8. Mita, K.; Tsuji, H.; Morimyo, M.; Takahashi, E.; Neno, M.; Ichimura, S.; Yamauchi, M.; Hongo, E.; Hayashi, A. (1995) The human gene encoding the largest subunit of RNA polymerase II. *Gene* 159: 285-286.
9. Pravtcheva, D.; Rabin, M.; Bartolomei, M.; Corden, J.; Ruddle, F. H. (1986) Chromosomal assignment of gene encoding the largest subunit of RNA polymerase II

in the mouse. *Somat. Cell Molec. Genet.* 12: 523-528.

13. Wintzerith, M., Acker, J., Vicaire, S., Vigneron, M. and C. Kedingner (1992) Complete sequence of the human RNA polymerase II largest subunit. *Nucleic Acids Res.* 20: 910.

5 10. Cornelis RS, van Vliet M, Vos CB, et al. (1994) Evidence for a gene on 17p13.3, distal to TP53, as a target for allele loss in breast tumors without p53 mutations. *Cancer Res.* 54:4200-6.

11. Lindblom A, Skoog L, Rotstein S, Werelius B, Larsson C, Nordenskjold M. (1993) Loss of heterozygosity in familial breast carcinomas. *Cancer Res.* 53:4356-61.

10 12. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* 51:5794-9.

13. Singh S, Simon M, Meybohm I, et al. (1993) Human breast cancer: frequent p53 allele loss and protein over expression. *Hum Genet.* 90:635-40.

15 14. Thorlacius S, Borresen AL, et al. (1993) Somatic p53 mutations in human breast carcinomas in an Icelandic population: a prognostic factor. *Cancer Res.* 53:1637-41.  
15. Tsuda H, Hirohashi S. (1994) Association among p53 gene mutation, nuclear accumulation of the p53 protein and aggressive phenotypes in breast cancer. *Int J Cancer.* 57:498-503.

20 16. Watatani M, Nagayama K, Imanishi Y, et al. (1993) Genetic alterations on chromosome 17 in human breast cancer: relationships to clinical features and DNA ploidy. *Breast Cancer Res Treat.* 28:231-9.

17. Chen LC, Neubauer A, Kurisu W, et al. (1991) Loss of heterozygosity on the short arm of chromosome 17 is associated with high proliferative capacity and DNA aneuploidy in primary human breast cancer. *Proc Natl Acad Sci USA.* 88:3847-51.

25 18. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* 51:5794-9.

19. Burner GC, Rabinovitch PS, Haggitt RC, et al. (1992) Neoplastic progression in ulcerative colitis: histology, DNA content, and loss of a p53 allele [see comments]. *Gastroenterology.* 103:1602-10.

20. Cunningham C, Dunlop MG, Wyllie AH, Bird CC. (1993) Deletion mapping in colorectal cancer of a putative tumour suppressor gene in 8p22-p21.3. *Oncogene*. 8:1391-6
21. Kikuchi-Yanoshita R, Konishi M, Ito S, et al. (1992) Genetic changes of both p53 alleles associated with the conversion from colorectal adenoma to early carcinoma in familial adenomatous polyposis and non-familial adenomatous polyposis patients. *Cancer Res*. 52:3965-71.
22. Yin J, Harpaz N, Tong Y, et al. (1993) p53 point mutations in dysplastic and cancerous ulcerative colitis lesions. *Gastroenterology*. 104:1633-9.
23. Iacopetta B, DiGrandi S, Dix B, et al. (1994) Loss of heterozygosity of tumour suppressor gene loci in human colorectal carcinoma. *Eur J Cancer*. 5:664-70.
24. Law DJ, Olschwang S, Monpezat JP, et al. (1988) Concerted nonsyntenic allelic loss in human colorectal carcinoma. *Science*. 241:961-5.
25. Lothe RA, Nakamura Y, Woodward S, Gedde DT, Jr., White R. (1988) VNTR (variable number of tandem repeats) markers show loss of chromosome 17p sequences in human colorectal carcinomas. *Cytogenet Cell Genet*. 48:167-9.
26. Foulkes WD, Stamp GW, Afzal S, et al. (1995) MDM2 over expression is rare in ovarian carcinoma irrespective of TP53 mutation status. *Br J Cancer*. 72:883-8.
27. Phillips NJ, Ziegler MR, Radford DM, et al. (1996) Allelic deletion on chromosome 17p13.3 in early ovarian cancer. *Cancer Res*. 56:606-11.
28. Foulkes WD, Black DM, Stamp GW, Solomon E, Trowsdale J. (1993) Very frequent loss of heterozygosity throughout chromosome 17 in sporadic ovarian carcinoma. *Int J Cancer*. 54:220-5.
29. Gallion HH, Powell DE, Morrow JK, et al. (1992) Molecular genetic changes in human epithelial ovarian malignancies [see comments]. *Gynecol Oncol*. 47:137-42.
30. Phillips N, Ziegler M, Saha B, Xynos F. (1993) Allelic loss on chromosome 17 in human ovarian cancer. *Int J Cancer*. 54:85-91.
31. Seruca R, David L, Castedo S, Veiga I, Borresen AL, Sobrinho-Simoes M. (1994) p53 alterations in gastric carcinoma: a study of 56 primary tumors and 204 nodal

metastases. *Cancer Genet Cytogenet.* 75:45-50.

32. Kim CJ, Kim WH, Kim CW, Lee JB, Lee CK, Kim YL. (1995) Detection of 17p loss in gastric carcinoma using polymerase chain reaction. *Lab Invest.* 72:232-6.

33. Ranzani GN, Renault B, Pellegata NS, et al. (1993) Loss of heterozygosity and K-ras gene mutations in gastric cancer. *Hum Genet.* 92:244-9.

34. Sano T, Tsujino T, Yoshida K, et al. (1991) Frequent loss of heterozygosity on chromosomes 1q, 5q, and 17p in human gastric carcinomas. *Cancer Res.* 51:2926-31.

35. Frankel RH, Bayona W, Koslow M, Newcomb EW. (1992) p53 mutations in human malignant gliomas: comparison of loss of heterozygosity with mutation frequency. *Cancer Res.* 52:1427-33.

36. Hermanson M, Funa K, Koopmann J, et al. (1996) Association of loss of heterozygosity on chromosome 17p with high platelet-derived growth factor alpha receptor expression in human malignant gliomas. *Cancer Res.* 56:164-71.

37. Aoki T, Mori T, Du X, Nisihira T, Matsubara T, Nakamura Y. (1994) Allelotype study of esophageal carcinoma. *Genes Chromosomes Cancer.* 10:177-82.

38. Tsuchiya E, Nakamura Y, Weng SY, et al. (1992) Allelotype of non-small cell lung carcinoma--comparison between loss of heterozygosity in squamous cell carcinoma and adenocarcinoma. *Cancer Res.* 52:2478-81.

39. Hiyama K, Ishioka S, Shirotani Y, et al. (1995) Alterations in telomeric repeat length in lung cancer are associated with loss of heterozygosity in p53 and Rb. *Oncogene.* 10:937-44.

#### **Example 26: TATA Associated Factor 30 kD subunit (TAF2H) - Target Gene VARIA 520**

*The human TAF2H gene encodes a component of the transcriptional apparatus*

Transcription initiation by RNA polymerase II requires the assembly of a complex of

basic transcription factors which include TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, TFIIG/TFIIJ and TFIIH/BTF2 into a preinitiation complex (1,2). TFIID is the first factor to contact the promotor, and subsequent assembly of the transcription complex is dependent on TFIID binding. TFIID is a 700-750 kD multiprotein complex which includes TATA binding protein (TBP) and between eight and 13 TBP-associated factors (TAFs) ranging from 250 to 17 kDa. The TAFs have been shown necessary to reconstitute activation of transcription *in vitro*, leading to the hypothesis that some TAFs link transcription activation domains to the basal transcription complex. The TFIID complex also supports transcription from TATA-less promoters, while TBP fails to do so. Therefore TAFs may also contribute to formation of stable initiation complexes by interacting directly with DNA (2). Conditional temperature sensitive Chinese hamster mutants of another TAF, TAFII250, were detected because, at the non-permissive temperature, DNA synthesis was inhibited leading to arrest of cell division at the G1 phase (3,4). Transfection of a human TAFII250 gene relieved the block at the non-permissive temperature. Thus an essential role has been proven for TAFs in mammalian cells.

A gene (TAF2H) encoding the 30 kDa human TAF protein (TAFII30) was cloned and its functional properties examined by Jacq, et al. (5). The protein was shown to be present in a subset of TFIID complexes and to mediate transcriptional activation by a specific region of the estrogen receptor. Estrogen mediated transcriptional activation could be abrogated by adding an antibody against TAFII30. TAFII30 was not required for basal transcription or for transcription activation by VP-16. It is likely that TAFII30 is required for transcriptional activation by a variety of other transactivating proteins, and is therefore essential for cell proliferation or cell survival.

*The human TAF2H gene and mRNA have sequence variants*

A human TAF2H cDNA has been cloned and sequenced (5). It encodes a cDNA of 756 nucleotides including a 5' untranslated region of 17 nucleotides, a 657 nucleotide

269

232/116

coding region specifying 218 amino acids, and an 82 nucleotide 3' untranslated region (GenBank accession U13991; see annotated TAF2H cDNA sequence). (Note that the numbering of the sequence in ref. 5 differs slightly from that in the GenBank accession.) We undertook a systematic search for DNA variance in the cDNA of TAF2H by analysing 36 unrelated individuals using the single strand conformation polymorphism technique. Primers were designed for amplification. SSCP analysis revealed 1 polymorphism, and subsequent DNA sequence analysis confirmed a G vs. A transition at nucleotide 554 (nt 556 of the sequence in ref. 3) of the coding sequence. This variance does not alter the protein coding sequence. Eight of 36 individuals surveyed are heterozygotes (22%). The variance occurs in North American Whites (3/16 = 19%), North American Blacks (2/4) and Hispanics (3/3).

*The human TAF2H gene maps to chromosome 11p15.5-p15.2* The human TAF2H cDNA has been mapped to 11p15.5-p15.2 by fluorescent *in situ* hybridization (6). There appears to be a single TAF2H locus. *Chromosome band 11p15-p14 is a site of frequent loss of heterozygosity*

There have been many studies of LOH on 11p, particularly the 11p15 and 11p13 segments where the Beckwith-Weidemann syndrome and WT1 genes reside. As a result there are many studies of LOH in 11p15.5, particularly focusing on breast, cervix, kidney, liver, lung, ovarian, stomach and testicular cancers. These studies show that the 11p15.5 band of chromosome 11 is frequently reduced to one copy (7-24). For example, LOH occurs in approximately 13-33% of breast cancers (7-9), 14-42% of cervical cancers (10), 0-50% of liver cancers (11,12), 0-80% of lung cancers (13-15), 18-54% of ovarian cancers (14,15), 0-71% of stomach cancers (18) and 0-50% of testicular cancers (19,20). Other studies show that 11p15.5 LOH may also be frequent in bladder cancer (21), esophageal cancer (22), some leukemias (23) and sarcomas (24). Many deletions in the 11p15.5 region span relatively short chromosomal segments (2 - 10 megabases; see ref. 13).



## References

1. Buratowski, S.(1994) The basics of basal transcription by RNA polymerase II. *Cell* 77: 1-3.

5 2. Tjian, R. and T. Maniatis (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77: 5-8.

3. Sekiguchi, T., Miyata, T. and T. Nishimoto (1988) Molecular cloning of the cDNA of human X chromosomal gene (CCG1) which complements the temperature sensitive G(1) mutants, tsBN462 and ts13, of the BHK cell line. *EMBO Journal* 7: 1683-1687.

10 4. Hisatake, K., Hasegawa, S., Takada, R., et al. (1993) The p250 subunit of native TATA box-binding factor TFIID is the cell -cycle regulatory protein CCG1. *Nature* 362: 172-181.

5. Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P. and L. Tora (1994) Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* 79: 107-117.

6. Scheer, E., Mattei, M.G., Jacq, X., Chambon, P. and L. Tora (1995) Organization and chromosomal localization of the gene (TAF2H) encoding the human TBP-associated factor II 30 (TAFII30). *Genomics* 29: 269-272.

7. Ali, I., Lidereau, R., Theilley, C. and R. Callahan (1987) Reduction to homozygosity of genes on chromosome 11 in human breast neoplasia. *Science* 238:185-8.

8. Winqvist, R., Mannermaa, A., Alavaikko, M., Blanco, G., Taskinen, P.J., Kiviniemi, H., Newsham, I. and W. Cavenee (1993) Refinement of regional loss of heterozygosity for chromosome 11p15.5 in human breast tumors. *Cancer Research* 53: 4486-4488.

9. Carter, S.L., Negrini, M., Baffa, R., et al. (1994) Loss of heterozygosity at 11q22-q23 in breast cancer. *Cancer Research* 54:6270-4.

10. Mitra, A.B., Murty, V.V.V.S., Li, R.G., et al. (1994) Allelotype analysis of cervical carcinoma. *Cancer Research* 54:4481.

11. Fujimori, M., Tokino, T., Hino, O., et al. (1991) Allelotype study of primary hepatocellular carcinoma. *Cancer Research* 51: 89-93.
12. Wang, H.P. and C.E. Rogler (1988) Deletions in human chromosome arms 11p and 13q in primary hepatocellular carcinomas. *Cytogenetics and Cell Genetics* 48:72-78.
13. Bepler, G. and Garcia-Blanco, M.A. (1994) Three Tumor Suppressor Regions on Chromosome 11p Identified by High Resolution Deletion Mapping in Human Non-Small Cell Lung Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 91:5513-7.
14. Iizuka, M., Sugiyama, Y., Shiraishi, M., Jones, C. and T. Sekiya (1995) Allelic losses in human chromosome 11 in lung cancers. *Genes, Chromosomes & Cancer* 13:40-46.
15. Weston, A., Willey, J.C., Modali, R., et al. (1989) Differential DNA sequence deletions from chromosomes 3, 11, 13 and 17 in squamous cell carcinoma, large-cell carcinoma and adenocarcinoma of the human lung. *Proc. Natl. Acad. Sci. U.S.A.* 86:5099-5103.
16. Kiechle-Schwartz, M., Bauknecht, T., Wienker, T., et al. (1993) Loss of Constitutional Heterozygosity on Chromosome 11p in Human Ovarian Cancer. *Cancer* 72:2423-32.
17. Viel, A., Giannini, F., Tumioti, L., Sopracordevole, F., Visentin, M.C. and M. Boiocchi (1992) Chromosomal localization of two putative 11p oncosuppressor genes involved in human ovarian tumors. *British Journal of Cancer* 66: 1030-1036.
18. Baffa, R., Negrini, M., Mandes, B., et al. (1996) Loss of heterozygosity for chromosome 11 in adenocarcinoma of the stomach. *Cancer Research* 56: 268-72.
19. Lothe, R.A., Hastie, N., Heimdal, K., et al. (1993) Frequent loss of 1p13 and 11p15 loci in male-germ cell tumors. *Genes, Chromosomes & Cancer* 7: 96-101.
20. Smith, R.C., and Rukstalis, D.B. (1995) Frequent Loss of Heterozygosity at 11p Loci in Testicular Cancer. *The Journal of Urology* 153: 1684-7.
21. Shaw, M.E. and Knowles, M.A. (1995) Deletion Mapping of Chromosome 11 in Carcinoma of the Bladder. *Genes, Chromosomes & Cancer* 13: 1-8.

272

232/116

22. Shibagaki, I., Shimada, Y., Wagata, T., Ikenaga, M., Imamura, M. and K. Ishizaki (1994) Allelotype analysis of esophageal squamous cell carcinoma. *Cancer Research* 54: 2996-3000.
23. Ahuja, H.G., Foti, A., Zhou, D.J. and M.J. Cline (1990) Analysis of proto-oncogenes in acute myeloid leukemia: loss of heterozygosity for the Ha-ras gene. *Blood* 75: 819-822.

### Example 27 - cDNA synthesis

In order to analyze an essential gene for sequence variances, it is generally useful to have a cDNA(s) containing the coding sequence for further sequencing or amplification purposes. cDNAs for some genes are available, however, in some cases it is useful to synthesize the cDNA *de novo*. Methods for obtaining cDNA are known to those skilled in the art, as are methods for sequencing or amplifying the cDNA or portions thereof. An example of a useful cDNA production protocol is provided below, however, as recognized by those skilled in the art, other specific protocols can also be used.

### cDNA Production

\*\* Make sure that all tubes and pipette tips are RNase-free. (Bake them overnight at 100oC in the vacuum oven to make them RNase-free.)

1 Add the following to a RNase-free 0.2 ml micro-amp tube and mix gently:

24 ul water (DEPC treated)

12 ul RNA (1ug/ul)

12 ul random hexamers(50 ng/ul)

2 Heat the mixture to 70oC for ten minutes.

3 Incubate on ice for 1 minute.

273

232/116

4 Add the following:

16 ul 5 X Synthesis Buffer

8 ul 0.1 M DTT

5 4 ul 10 mM dNTP mix (10 mM each dNTP)

4 ul SuperScript RT II enzyme

Pipette gently to mix.

5 Incubate at 42°C for 50 minutes.

6 Heat to 70°C for ten minutes to kill the enzyme, then place it on ice.

10 7 Add 160 ul of water to the reaction so that the final volume is 240 ul.

8 Use PCR to check the quality of the cDNA. Use primer pairs that will give a  
~800 base pair long piece. See "PCR Optimization" for the PCR protocol.

15 The following chart shows the reagent amounts for a 20 ul reaction, a 80 ul  
reaction, and a batch of 39 (which makes enough mix for 36) reactions:

	20 ul X 1 tube	80 ul X 1 tube	80ul X 39 tubes	
water	6 ul	24 ul	936	water
RNA	3 ul	12 ul		RNA
random hexamers	3 ul	12 ul	468	random hexamers
synthesis buffer	4 ul	16 ul	624	synthesis buffer
0.1 M DTT	2 ul	8 ul	312	0.1 M DTT
10mM dNTP	1 ul	4 ul	156	10mM dNTP
SSRT	1 ul	4 ul	156	SSRT

30

#### Example 28 - Variance detection by SSCP

This example describes the SSCP technique as used for the identification of sequence variances of the exemplary genes, which were then sequenced to confirm the specific base variances. One common technique currently employed in the identification of such single nucleotide differences is the single strand conformation polymorphism (SSCP) method. (originally described in Orita, *et al.*, "Rapid and Sensitive Detection of Point Mutations and DNA Polymorphisms Using the Polymerase Chain Reaction, *Genomics*, 5:874-879 (1989)) Also employed are restriction fragment length polymorphism (RFLP), heteroduplex analysis, ligase chain reaction (LCR), denaturing gradient gel electrophoresis (DGGE) (Myers, Maniatis, and Lerman, *Methods Enzymol.*, 155:501-527 (1987)) or direct nucleotide sequencing. A review of polymorphism detection techniques, including SSCP, is provided in Grompe, 1993, *Nature Genetics* 5:111-117, which includes a comparison of the commonly used methods.

The SSCP method reveals the presence of sequence variation between individuals as shifts in electrophoretic mobility, but does not show the sequence itself. Direct sequencing of DNAs with altered mobility in the SSCP assay identifies the precise nucleic acid sequence differences among the various alleles. From the nucleic acid sequence data, the amino acid sequence can be determined. One example of the use of this technique is in Pelletier *et al.*, *Cell*, 67:437-447 (1991). The single strand conformation polymorphism methodology is effective for scanning essential genes for sequence variants. It remains the standard technique in human genetics for variance detection, with numerous studies of its efficacy (>90%) and schemes for improved throughput. The SSCP method has been shown to be quite sensitive in the detection of single base changes, for example as shown in Ravnik-Glava *et al.*, 1994, *Human Mol. Genet.* 3:801-807 (human cystic fibrosis gene) and Glava & Dean, 1993, *Human Mutation* 2:404-414 (mouse  $\beta$ -globin gene).

A flow chart of the SSCP method as used to identify essential gene sequence variants is shown in Fig. 2 (SSCP OVERVIEW). The method involves the steps of 1) PCR

275

232/116

amplifying a portion of an essential gene cDNA of known sequence (labeled products),  
2) selecting restriction enzymes which will produce fragments approximately 100-400  
bases in length for 3 independent digestions of the PCR products, 3) heat denaturing  
the digestion products, 4) running single strand digestion products on non-denaturing  
5 gels, 5) identifying bands having different mobilities when compared between  
individuals, thereby identifying potential sequence variants, 6) sequence at least the  
region around the potential sequence variance, that region being identified by  
comparison of the expected fragment sizes resulting from the digestions, 7) record the  
specific location and base identity of the confirmed sequence variant, 8) calculate the  
10 percent occurrence of each sequence variance for the gene as found for the sample of  
the population. The method is further described in Example 2.

Single strand conformation polymorphism screening is a widely used technique for  
identifying an discriminating DNA fragments which differ from each other by as little  
15 as a single nucleotide. As originally developed by Orita (supra), the technique was  
used on genomic DNA, however the same group showed that the technique works very  
well on PCR amplified DNA as well. In the last 8 years the technique has been used  
in hundreds of published papers, and the modifications of the technique have been  
described in dozens of papers. The enduring popularity of the technique is due to (1)  
20 a high degree of sensitivity to single base differences (>90%) (2) a high degree of  
selectivity, measured as a low frequency of false positives, and (3) technical ease.  
SSCP is almost always used together with DNA sequencing because SSCP does not  
directly provide the sequence basis of differential fragment mobility. The basic steps  
of the SSCP procedure are described below and summarized in Fig. 2 in flow chart  
25 form.

Because the intent of our SSCP screening was to identify as many target gene  
variances as practically possible, we developed a protocol designed to look at a  
relatively large number of individuals (36) with a high degree of redundancy, so as to  
minimize both the false negative and false positive rates.

The 36 individuals examined are reasonably representative of most of the worlds major populations. The racial or geographic origin of the 36 cell lines is detailed in the Target Summary Tables (Figure 5). All cell lines are EBV immortalized lymphoblastoid cells obtained from the Coriell Cell Repository (Camden, NJ), which includes the racial/ethnic/geographic background of cell line donors in its catalog. The cell lines were also selected for their rapid growth rates. In several cases a panel of cDNAs isolated from French Canadians was used instead, or in addition to, the Coriell panel.

SSCP was used to analyze cDNAs (rather than genomic DNAs) because in many cases the full genomic sequence of the target gene is not available, however, the technique is also applicable to genomic sequences. To produce cDNA requires RNA. Therefore each of the 36 cell lines was grown to mass culture and RNA was isolated using the acid/phenol protocol, sold in kit form as TRIAZOL™ by Life Technologies (Gaithersburg, MD). The unfractionated RNA was used to produce cDNA by the action of a modified Maloney Murine Leukemia Virus Reverse Transcriptase, purchased in kit form from Life Technologies (SUPERScript II™ kit). The reverse transcriptase was primed with random hexamer primers to initiate cDNA synthesis along the whole length of the RNAs. This proved useful later in obtaining good PCR products from the 5' ends of some genes.

Material for SSCP analysis was prepared by PCR amplification of the cDNA in the presence of one <sup>32</sup>P labeled dNTP (usually <sup>32</sup>P dCTP). Usually the concentration of nonradioactive dCTP was dropped from 200 uM (the standard concentration for all four dNTPs) to about 100 uM, and <sup>32</sup>P dCTP was added to a concentration of about 0.1-0.3 uM. This involved adding a 0.3- 1 ul (3-10 uCi) of <sup>32</sup>P dCTP to a 10 ul PCR reaction. All radioactivity was purchased from DuPont/New England Nuclear.

The customary practice is to amplify about 200 base pair PCR products for SSCP, however, we found that it was preferable to amplify about 0.8-1.4 kb fragments and

277

232/116

then use several cocktails of restriction endonucleases to digest those into smaller fragments of about 0.1-0.4kb, aiming to have as many fragments as possible between .15 and .3 kb. The digestion strategy had the advantage that less PCR was required, reducing both time and costs. Also, we routinely performed three different digests on each sample (for all 36 cDNAs), and then ran each of the digests separately on SSCP gels. This had the effect of increasing the redundancy of our method, lessening both the false negative and false positive rates. For example: a site of variance might lie within 2 bases of the end of a fragment in one digest, and as a result not affect the conformation of that strand; the same variance, in a second or third digest, would likely lie in a location more prone to affect strand folding, and therefore be detected by SSCP.

After digestion, the radiolabeled PCR products were diluted 1:5 by adding formamide load buffer (80% formamide, 1X SSCP gel buffer) and then denatured by heating to 90°C for 10 minutes, and then allowed to renature by quickly chilling on ice. This procedure (both the dilution and the quick chilling) promotes intra- (rather than inter-) strand association and secondary structure formation. The secondary structure of the single strands influences their mobility on nondenaturing gels, presumably by influencing the number of collisions between the molecule and the gel matrix (i.e., gel sieving). Even single base differences consistently produce changes in intrastrand folding sufficient to register as mobility differences on SSCP.

The single strands were then resolved on two gels, one a 5.5% acrylamide, 0.5X TBE gel, the other an 8% acrylamide, 10% glycerol, 1X TTE gel. The use of two gels provides a greater opportunity to recognize mobility differences. Both glycerol and acrylamide concentration have been shown to influence SSCP performance. The gel apparatus we use (from Owl Scientific, MA) allows 108 samples to be loaded per gel. Since all 36 samples are routinely digested with three different endonuclease mixes there are 108 samples to be analyzed for each PCR product. By routinely analyzing three different digests under two gel conditions (effectively 6 conditions), and by



looking at both strands under all 6 conditions, we achieve a 12-fold sampling of each base pair of cDNA.

5 All of the sequence variances described in this disclosure were determined by DNA cycle sequencing of  $^{32}\text{P}$  labeled PCR products using the femtomole DNA cycle sequencing kit from Promega (WI) and the instructions provided with the kit. Fragments were selected for DNA sequencing based on their behavior in the SSCP assay.

10 **Example 29 - Variance detection by using T4 endonuclease VII mismatch cleavage method**

15 The enzyme T4 endonuclease VII is derived from the bacteriophage T4. T4 endonuclease VII is used by the bacteriophage to cleave branched DNA intermediates which form during replication so the DNA can be processed and packaged. T4 endonuclease can also recognize and cleave heteroduplex DNA containing single base mismatches as well as deletions and insertions. This activity of the T4 endonuclease VII enzyme can be exploited to detect sequence variances present in the general  
20 population.

The following are the major steps involved in identifying sequence variations in a candidate gene by T4 endonuclease VII mismatch cleavage:

- 25
1. Amplification by the polymerase chain reaction (PCR) of 400-600 bp regions of the candidate gene from a panel of DNA samples. The DNA samples can either be cDNA or genomic DNA and will represent some cross section of the world population.
  2. Mixing of a fluorescently labeled probe DNA with the sample DNA. Heating

279

232/116

and cooling the mixtures causing heteroduplex formation between the probe DNA and the sample DNA.

3. Addition of T4 endonuclease VII to the heteroduplex DNA samples. T4 endonuclease will recognize and cleave at sequence variance mismatches formed in the heteroduplex DNA.
4. Electrophoresis of the cleaved fragments on an ABI sequencer to determine the site of cleavage.
5. Sequencing of a subset of PCR fragments identified by T4 endonuclease VI to contain variances to establish the specific base variation at that location.

A more detailed description of the procedure is as follows:

A candidate gene sequence is downloaded from an appropriate database. Primers for PCR amplification are designed which will result in the target sequence being divided into amplification products of between 400 and 600 bp. There will be a minimum of a 50 bp of overlap not including the primer sequences between the 5' and 3' ends of adjacent fragments to ensure the detection of variances which are located close to one of the primers.

Optimal PCR conditions for each of the primer pairs is determined experimentally. Parameters including but not limited to annealing temperature, pH,  $MgCl_2$  concentration, and KCl concentration will be varied until conditions for optimal PCR amplification are established. The PCR conditions derived for each primer pair is then used to amplify a panel of DNA samples (cDNA or genomic DNA) which is chosen to best represent the various ethnic backgrounds of the world population or some designated subset of that population.

One of the DNA samples is chosen to be used as a probe. The same PCR conditions used to amplify the panel are used to amplify the probe DNA. However, a

280

232/116

fluorescently labeled nucleotide is included in the deoxy-nucleotide mix so that a percentage of the incorporated nucleotides will be fluorescently labeled.

5 The labeled probe is mixed with the corresponding PCR products from each of the DNA samples and then heated and cooled rapidly. This allows the formation of heteroduplexes between the probe and the PCR fragments from each of the DNA samples. T4 endonuclease VII is added directly to these reactions and allowed to incubate for 30 min. at 37 C. 10 ul of the Formamide loading buffer is added directly to each of the samples and then denatured by heating and cooling. A portion of each  
10 of these samples is electrophoresed on an ABI 377 sequencer. If there is a sequence variance between the probe DNA and the sample DNA a mismatch will be present in the heteroduplex fragment formed. The enzyme T4 endonuclease VII will recognize the mismatch and cleave at the site of the mismatch. This will result in the appearance of two peaks corresponding to the two cleavage products when run on the ABI 377  
15 sequencer.

Fragments identified as containing sequencing variances are subsequently sequenced using conventional methods to establish the exact location and sequence variance.

### 20 **Example 30 - Identification of Sequence Variances by Informatics-based analysis of gene-sequence databases**

In addition to and/or in conjunction with the molecular biology based approaches for identifying sequence variances in genes, particularly in essential genes, such sequence  
25 variances can be identified by analysis of public and/or private genetic sequence databases. Such information can be either genomic or cDNA sequence information.

The data base analysis process includes the following major steps:

1.

1. capture of homologous sequences of a particular gene from data bases. It is preferable to obtain a large number of independent sequences of a particular gene

5

2. analysis of collected sequences of a particular gene to identify authentic sequence variances. This step involves the discrimination of authentic sequence variances, which are sequence variances which actually exist in the population, from sequencing errors and artifacts. It is expected that about 0.1-0.3% of the bases will occur as true variances, while the frequency of sequencing artifacts is expected to be 1-3%. This discrimination utilizes the expected frequencies of occurrence of specific types of nucleotide sequence changes. Such information includes the characteristic frequency of specific transitions and transversions and of the characteristic frequency of deletions and insertions in authentic variations. It uses the frequency of occurrence of known types of sequencing artifacts such as single base insertions or deletions adjacent to repeated C or G nucleotides. Additional information for such discrimination is provided if particular putative authentic variations are observed in multiple independently derived sequences of the gene.

10

15

20

25

An implementation of this sequence variance identification process utilizes a reference sequence of an essential gene. Preferably, the reference sequence is a high quality sequence, meaning that there is a low frequency of occurrence of sequencing errors or artifacts. The second step is the retrieval of allelic sequences of that essential gene from available databases such as the BLAST server, the UNIGENE database, or other such sequence database. Such allelic sequences need not be complete, but are preferably long enough to ensure that they are in fact allelic sequences. The third step involves alignment analysis to identify and tabulate sequence differences between the different available sequences. An algorithm for such analysis is the Smith-Waterman local alignment algorithm. Use of an algorithm of this type involves a series of pair-

wise alignments of each retrieved sequence with the reference sequence. The fourth step involves analysis of the observed sequence differences and assignment of a probability that each sequence difference represents an authentic variance. This analysis utilizes program filters which are combined in a weighted fashion to determine a final probability. Such program filters include comparison of the observed difference with common mutational changes and sequencing errors, a weighting of the reliability of a particular retrieved sequence based on the total number of differences observed, a weighting based on the location within a retrieved sequence where a change was observed and a significant weighting based on the observance of a particular difference in multiple independently derived retrieved sequences.

Using such an implementation, a database analysis with respect to a particular reference sequence produces a list of putative authentic sequence variances and a probability for each of those variances that the sequence difference is an authentic variance. As described above, the probability is obtained through the use of a series of weighted program filters and thus these filters are modified to produce optimal authentic variance discrimination.

#### **Example 31 - Antiproliferative effects of variance specific inhibition of RPA70**

This example describes experiments showing the practicality and utility of variance-specific inhibition of essential genes for cancer therapy. Specifically, this example describes *in vitro* experiments showing the design and production of variance-specific oligonucleotides for antisense inhibition of variant alleles of the essential Replication Protein A, 70 kDa subunit (RPA70) for inhibition of RPA70 mRNA, and the use of these oligonucleotides to inhibit cell proliferation and to reduce the number of cells in a variance-specific manner.

*Variance-specific inhibition and cell killing with antisense oligonucleotides against*

*RPA70*

These experiments with RPA70 illustrate the feasibility of each of the steps for development of a variance specific inhibitor:

5       Select candidate target gene essential for cell survival or proliferation. As described above, RPA is essential for replication in prokaryotic and eukaryotic cells, mitochondria, phage, viruses and in *in vitro* (SV40) replication systems. The protein is a heterotrimer required for loading DNA polymerase onto the DNA template during cell replication. The 70 kDa subunit, RPA70, is a single strand binding protein that  
10       mediates the interaction of RPA with DNA. Without this protein, the replication complex does not associate with DNA and the replication of DNA does not occur.

Confirm chromosome location and LOH frequency. RPA70 is encoded by a single gene locus on chromosome 17p13.3, immediately adjacent to the p53 gene at 17p13.1.  
15       LOH involving chromosome band 17p13.3 has been documented in 50-70% of colon, lung, breast, and ovarian cancers. LOH at this locus also occurs in other cancers. The inventor as confirmed LOH involving RPA 70 in breast, colon, lung and other cancers.

Identify common variances in the normal population. We have identified five common  
20       variances in the RPA70 gene (Figure 8). The most common occurs in 42% of the normal population. One variance alters the amino acid sequence and is present in 25% of the normal population (44% of Caucasians). This variance occurs within the active DNA binding domain (discussed below). These variances are described in the description above and in Fig. 1.

25       Demonstrate antiproliferative effects due to inhibition of candidate gene. The inventor has shown that inhibition of RPA70 in T24 bladder carcinoma cells with an antisense oligonucleotide reduces cell number. This effect is comparable to treatment of these cells with antisense oligonucleotide against *ras*, previously shown to have antitumor

284

232/116

effects *in vitro* and *in vivo* (Figure 9).

Design variance-specific inhibitor. Variance specific antisense oligonucleotides were designed to differentially inhibit the two variant forms of RPA70. Experiments were performed using tumor cell lines that are homozygous for each form of the target gene. Figure 10 shows inhibition of mRNA levels in Mia Paca II cells by the 13085 oligonucleotide which matches the variance in these cells. In contrast, in T24 cells (and A549 cells, see below) the 12781 oligonucleotide matches the target gene and inhibits mRNA levels. In both cell lines neither the control oligonucleotide differing by one base (13085 in T24 cells and 12781 in Mia Paca II cells) nor a random-sequence oligonucleotide control (13706) inhibit mRNA levels to the same extent as the matched oligonucleotide.

Figure 10 demonstrates that the RPA 70 mRNA can be specifically down regulated in an allele-specific manner. However, the 13085 oligomer used also has a small effect on the level of the unmatched RNA. In order to increase the discrimination we altered the structure of the targeting oligomer, 13085. The results are shown in Figure 11. By shortening the oligomer we retain its ability to down-regulate its matched target RNA (Mia Paca II cells, right half of Figure 11). Strikingly, however, this alteration dramatically altered the ability of this oligomer to down-regulate the mismatched variant RNA T24 cells, left half of Figure 11. The reciprocal regulation by oligomer 12781 was augmented by altering transfection conditions. These data suggest that even simple changes to the rudimentary "first generation" chemistry and transfection techniques can have significant effects in enhancing the ability of the oligomers to recognize and down regulate specific mRNAs.

Achieve variance-specific antiproliferative effects in cancer cells. Cell proliferation in each cell line, determined by BrdU incorporation, was suppressed to a greater degree by the matched oligonucleotide than by the controls differing by one base (Figure 12).

285

232/116

Cell proliferation in A549 cells was inhibited by oligomer 12781 to a greater degree than by oligomer 13085. Cell proliferation in Mia Paca 11 cells was inhibited more by oligomer 13085.

5 Additional studies were performed to characterize the antiproliferative effect in A549 cells (12781 genotype). A dose response curve demonstrates inhibition of BrdU incorporation by the matched oligonucleotide (12781) at concentrations 8-fold lower than the oligonucleotide with one base mismatch (13085) (Figure 13).

10 Cell survival was measured by staining cells with Sulforhodamine B dye 72 hours after treatment with oligonucleotides. Dose dependent reductions in cell number were observed in cells treated with the matched oligonucleotide (12781) but not with an oligonucleotide containing the one base mismatch (13085) (Figure 14). In contrast, in  
15 Mia Paca II cells, more cell killing was observed with the 13085 oligonucleotide than with the 12781 oligonucleotide (Figure 15). The oligonucleotides used in these studies have not been optimized for achieving allele-specific effects. Oligonucleotides using advanced chemistries can be utilized to optimize the potency and provide greater discrimination between variant targets at lower levels.

20

#### **Example 32 - variance specific inhibition of essential genes**

This example describes experiments showing the practicality and utility of variance-specific inhibition of essential genes for cancer therapy including RNA Pol II, and ribonucleotide reductase. Specifically, this example describes *in vitro* experiments  
25 showing the design and production of variance-specific oligonucleotides for antisense inhibition of variant alleles of the essential Ribonucleotide Reductase (RR), the design and production of variance-specific oligonucleotides against RR, and the use of these oligonucleotides to inhibit RR mRNA in a variance-specific manner.



*Variance-specific inhibition of Ribonucleotide Reductase.*

Ribonucleotide Reductase (RR) is an essential gene of nucleoside metabolism. Inhibitors of this function are known to be cell lethal. Two variances were discovered at position 2410 and 2419. Oligonucleotides were synthesized to a sequence spanning these two variations. In one case the oligomer targeted the GnnnnnnnnA variation (oligomer Varia 2410GA or RR2410GA) and in the other case the oligomer targeted the AnnnnnnnnG variant (oligomer Varia 2410AG or RR2410AG). In Mia Paca II cells which contain the GnnnnnnnnA variance, the RR2410GA antisense oligomer dramatically knocked down the level of RR mRNA. However, the oligomer targeting the other variance, oligomer Varia 2410AG, had little to no effect on the level of mRNA (Figure 16). The reciprocal regulation was demonstrated in MDA-MB 468 cells which express the other variance, AnnnnnnnnG (Figure 17). In these cells Varia 2410AG dramatically lowered the level of RR mRNA. In contrast, Varia 2410GA had no effect on the level of mRNA. These data taken together, are another example of allele-specific targeting of gene expression. We are also determining the effect of down regulating RR gene expression on cellular growth.

**Example 33 - variance specific inhibition of essential genes using advanced oligonucleotide chemistries.**

This example describes experiments showing the practicality and utility of variance-specific inhibition of essential genes for cancer therapy. Specifically, this example describes *in vitro* experiments showing the design and production of variance-specific oligonucleotides for antisense inhibition of variant alleles of the essential Glutamyl/prolyl tRNA Synthetase (EPRS), the design and production of variance-specific oligonucleotides against EPRS, and the use of these oligonucleotides to inhibit EPRS mRNA in a variance-specific manner.

287

232/116

Glutamyl-prolyl-tRNA synthetase (EPRS) is an essential gene, required for the synthesis of both glutamic acid tRNA and proline tRNA. Without EPRS protein synthesis is blocked. Two variances were discovered in this gene at positions 2963 and 2969 in the cDNA. We have demonstrated variance-specific inhibition of this gene with antisense oligonucleotides exploiting several different types of chemistry.

The experiments described above with RPA70 and RR utilized phosphorothioate chemistry. This chemistry was developed to achieve greater stability *in vivo*, and this compound has been used in several successful clinical trials. Phosphorothioates, however have low affinity for the RNA target, and, consequently, relatively lower specificity. We have achieved improved variance-specific inhibition using alternative chemistries. Specifically, we have synthesized hybrid oligonucleotides that contain both phosphorothioate and nucleotides with higher affinities. These hybrids contain "wings" consisting of six nucleotides with a 2' sugar modification (ethoxy-methoxy radical at the 2' position) and either a phosphorothioate or phosphodiester backbone. Between the "wings" is a 8 nucleotide sequence of phosphorothioates that overlaps the variance. (In these constructs the 5' position of cytosine has been methylated.) As shown in Figure 18, variance specific inhibition is observed with the conventional phosphorothioates. Greater inhibition of target mRNA is observed using the hybrid chemistries at lower doses. Inhibition by the matched hybrid oligomer, 14977, occurs at approximately 50-100 nM. The effect is extremely oligomer-specific. The mismatched oligomer, 14971, has no effect on mRNA levels at concentrations as high as 400 nM (Figure 19).

#### Example 34 - *in vivo* cancer therapy using oligonucleotides

This example describes reported *in vitro* and *in vivo* data on the treatment of cancer in animal models using antisense oligonucleotides against c-raf, showing the expected

correlation between *in vitro* suppression of mRNA and cell proliferation with oligonucleotides, and *in vivo* anticancer activity.

*In vitro* evidence for inhibition of mRNA by antisense oligonucleotides and inhibition of cell proliferation is commonly used to predict *in vivo* effects on tumors. This is exemplified by the publication by Monia et al (Nature Medicine, Volume 2 Number 6, June 1996) who demonstrated anticancer effects using oligonucleotides against C-raf kinase. *In vitro* treatment of human tumor cells with appropriate phosphorothioate antisense oligomers led to specific inhibition of C-raf kinase gene expression and subsequent decrease in cellular proliferation, IC<sub>50</sub>=50-100nM. Administration of C-raf antisense oligomers to nude mice having a tumor burden derived from these cells significantly inhibited tumor growth *in vivo*, IC<sub>50</sub>= 0.06-0.6 mg/kg. Remarkably, the investigators were able to show that the anti-C-raf oligomers down-regulated the level of C-raf kinase mRNA *in vivo* by assaying mRNA levels in cells removed from the tumor.

#### **Example 35 - *in vivo* cancer therapy by oligonucleotide inhibition of ras**

This example describes reported *in vivo* data showing an anticancer effect using an allele-specific inhibitor for suppression of mutant H-ras. Schwab *et al* (*Proc. Nat. Acad. Sci. USA* 91:10460-464, Oct 1994) demonstrated antitumor effects of an antisense oligonucleotide specific for the mutant ras in animal models. In these experiments HBL100 cells were transformed with the RAS oncogene. *In vitro* studies demonstrated that the RAS mRNA could be specifically down-regulated by a nanoparticle conjugated phosphodiester antisense oligomer. Only the transforming RAS mRNA was targeted by the oligomer. The normal cellular RAS mRNA, differing by a single base, was not affected by the antisense oligomer. The decrease in RAS expression was associated with a decrease in the growth rate of the cells. The

transformed HBL100 cells were injected into nude mice to form tumors; following subcutaneous injection of nanoparticle-conjugated phosphodiester antisense oligomers, Schwab et al measured both a decrease in targeted tumor weight and volume. Specificity for tumor cell growth correlated well with the *in vitro* data having a 5-fold differential between antisense and control groups.

The authors of this paper are proceeding with clinical trial of these oligonucleotides for the treatment of cancer, demonstrating the potential clinical utility of these methods.

#### Example 36. Variance detection by DGGE

This example describes denaturing gradient gel electrophoresis (DGGE), a technique used for the identification of DNA sequence variances in genomic DNA, cDNA or in PCR products amplified from genomic DNA or cDNA. The DGGE method was originally described by Fischer and Lerman (Two Dimensional Electrophoretic Separation of Restriction Enzyme Fragments of DNA. Methods in Enzymology, vol. 68: 183-191, 1979; DNA Fragments Differing by Single Base-Pair Substitutions are Separated in Denaturing Gradient Gels: Correspondence with Melting Theory. Proc. Natl. Acad. Sci. U.S.A. 80:1579, 1983) and has been improved since then by many investigators. See, for example: Myers, et al., Mutation Detection by PCR, GC-Clamps, and Denaturing Gradient Gel Electrophoresis, pp. 71-88 in Erlich, H.A., editor: PCR Technology: Principles and Applications for DNA Amplification, Stockton Press, New York, 1989; Myers, et al., Detecting Changes in DNA: Ribonuclease Cleavage and Denaturing Gradient Gel Electrophoresis, in Davies, K.E., editor: Genomic Analysis: A Practical Approach, IRL Press Ltd., Oxford, 1988, pp. 95-139; E.S. Abrams and V.P. Stanton Jr., Use of Denaturing Gradient Gel Electrophoresis, pp. 71-104 in Lilley, D.M.J. and Dahlberg, J.E., editors: DNA Structures. Part B: Chemical and Electrophoretic Analysis of DNA, Methods in

Enzymology, volume 212, Academic Press, 1992; .) Descriptions of current applications of the technique can be found in

5 The basic principal of DGGE involves the creation of a gradient of denaturant in a gel, which is then used to resolve double stranded DNA (or RNA) fragments on the basis of conformational differences associated with strand melting. The denaturant can be chemical (as in DGGE, where a gradient of formamide and urea is typically used) or thermal (as in a related technique called thermal gradient gel electrophoresis, or TGGE, where a gradient of heat is used). To obtain conditions where double stranded DNA  
10 is close to melting, DGGE gels are immersed in a heated bath of electrophoresis buffer, while TGGE gels have a fixed concentration of chemical denaturant.

15 As a double stranded DNA molecule migrates through a DGGE gel from a low concentration of denaturant at the origin to higher concentrations of denaturant toward the end of the gel it eventually reaches a level of denaturant that will cause partial melting. (Some design of DNA molecules is often necessary to assure that the partial melting will occur as desired; see below.) The concentration of denaturant required to melt a given DNA segment is highly sensitive to sequence differences in the DNA, including changes as subtle as a single nucleotide substitution. Partially melted DNA  
20 fragments move through gels at a much slower rates than their fully duplex counterparts. Thus two DNA fragments differing at a single nucleotide can be distinguished on the basis of their gel position after an appropriate period of electrophoresis: the fragment with the more stable structure (resulting from, for example, a G:C base pair in place of an A:T pair) will travel further in the gel than its  
25 less stable counterpart, because it will encounter the concentration of gradient required to melt it (and consequently dramatically retard or nearly stop its movement) at a point further along in the gel.

The DGGE method reveals the presence of sequence variation between individuals as

shifts in electrophoretic mobility, but does not show the sequence itself. Direct sequencing of DNA fragments (from different individuals) with altered mobility in the DGGE assay will reveal the precise sequence differences among them (see example 37, Variance Detection by DNA Sequencing). From the nucleic acid sequence data, the amino acid sequence can be determined and any amino acid differences can be identified.

The DGGE method is suitable for analysis of restriction enzyme digested genomic DNAs, as initially described by Lerman and co-workers (*supra*) and later extended (Gray, M. Detection of DNA Sequence Polymorphisms in Human Genomic DNA by Denaturing Gradient Blots, American Journal of Human Genetics, 50: 331-346, 1992). DGGE is equally suitable for analysis of cloned DNA fragments or DNA fragments produced by PCR. The analysis of cloned fragments or PCR fragments has the advantage that non-natural sequences, rich in G and C nucleotides can easily be added to the 5' ends (either flanking the cloning site or at the 5' ends of PCR primers). Such DNA fragments have very stable double stranded segments, called GC clamps, at one or both ends. The GC clamps alter the melting properties of the fragments, and can be designed so as to insure melting of the inter-primer segment of the PCR product at a lower temperature than the clamps, thereby optimizing the detection of sequence differences (see Myers *et alia*, *supra* and Myers *et alia*, Nearly All Single Base Substitutions in DNA Fragments Joined to a GC Clamp Can be Detected by Denaturing Gradient Gel Electrophoresis. Nucleic Acids Research 13: 3131, 1985). GC clamps can be rationally designed for any specific DNA fragment of known sequence by use of a computer program (MELT87, written by L. Lerman) that accurately predicts melting behavior based on analysis of primary sequence. When GC clamps are used correctly, the DGGE method is highly efficient at detecting DNA sequence differences. Not only are nearly 100% of differences detected, but the false positive rate is essentially zero. (Abrams, E.S., et alia, Comprehensive Detection of Single Base Changes in Human Genomic DNA Using Denaturing Gradient Gel

Electrophoresis and a GC Clamp. Genomics 7: 463-475, 1990.) Recently methods for increasing the throughput of DGGE have been developed, based on multiplex PCR.

The steps in carrying out DGGE with GC clamps are:

5

1. *Design DNA fragments with optimal melting behavior.* Select oligonucleotide primers, using GC clamps as necessary, to produce a single melting domain over the length of the sequence to be analyzed. (It may be necessary to divide the sequence into overlapping fragments to achieve this goal.) Design of primers and simulated analysis of fragments can be performed with the computer program described by Lerman. (Lerman, L.S. and Silverstein, K. Computational Simulation of DNA Melting and its Application to Denaturing Gradient Gel Electrophoresis. Methods in Enzymology 155: 482-501, 1987.) The output of the program is the melting map of the fragment, from which it will also be possible to determine the optimal range of denaturant in the gradient and the approximate electrophoresis time for fragments to reach the point of melting in the gradient.

10

15

2. *Amplify the fragment by PCR.* Procedures for optimizing PCR are briefly described in other examples and are well known in the art. Template DNA samples can either be cDNA or genomic DNA and will typically be drawn from a panel of unrelated individuals.

20

3. *Pour a denaturing gradient gel.* Briefly, make up two gel solutions containing the desired beginning and end concentrations of denaturant. The gel solutions are generally made up by mixing "0%" and "100%" denaturant stock solutions, where the 0% stock consists of 7% acrylamide in Tris-acetate EDTA (TAE) electrophoresis buffer, and the 100% stock is also 7% acrylamide in TAE, plus 40% formamide by volume and 7 molar urea. Equal volumes of the two solutions (e.g. twelve milliliters of each solution) are poured into the two chambers of a gradient maker (usually between 20 and 40% denaturant in the upstream chamber and 60 to 80% in the lower

25

one) immediately after addition of ammonium persulfate and TEMED for acrylamide polymerization. Open the stopcock of the gradient maker and pour the gradient gel. Usually gels are .75 to 1 mm in thickness, and gel combs that form 10-30 wells are used. With commercially available apparatus multiple gradient gels can be poured simultaneously. Suitable apparatus is sold by several vendors, including the BioRad (Hercules, CA) Dcode system and the C.B.S. Scientific DGGE system.

4. *Place the gel in a heated bath of electrophoresis buffer.* Gels are electrophoresed at elevated temperature which, together with the denaturant, brings the DNA fragments to their melting point. Gels are often run at 60°C in 1X TAE buffer, with constant recirculation of buffer to the upper buffer chamber. Once the gel has been placed in the heated tank and allowed to equilibrate it can be loaded. Multiple gels can be run simultaneously in the same tank with the apparatus listed above.

5. *Load and run gel.* Usually enough PCR product from each sample is loaded on the gel so that samples can be detected by a simple DNA staining procedure; use of radioactivity, dyes or hybridization procedures can thereby be avoided. At least 100 mg of each sample should be loaded, but preferably over 200 ng. Gel running conditions can be estimated from the output of the MELT87 program, however empirical adjustment will often be necessary. Usually a voltage of ~80 to 200V is applied for periods of 5-20 hours, depending on the characteristics of the fragments being analyzed.

6. *Stain and analyze gel.* After electrophoresis gels are stained with ethidium bromide, SYBR Green, silver or some other procedure. The location of PCR products produced with the same primer pairs should be compared. Altered location, and usually the appearance of two or more bands instead of one, signify the presence of DNA sequence differences. (The reason for more than two bands from a diploid sample is that during the terminal cycle of heating and cooling of the PCR



step heteroduplexes are formed between the maternally and paternally inherited alleles. If those alleles differ in sequence, the heteroduplexes will have mispaired nucleotides at the sites of difference. As a result the heteroduplexes will be less stable than either of the homoduplex species, and will consequently melt and be retarded in the gel at a lower concentration of denaturant. Altogether one may see four bands in such samples: two reciprocal heteroduplexes and two homoduplexes.) The specific pattern of fragments in each lane constitutes a signature for a specific nucleotide change.

7. *Sequence DNA fragments with altered mobility.* Examples of all different signatures should next be analyzed by DNA sequencing to identify the base difference(s) accounting for altered mobility in the gradient gel. See example 37 for a description of this procedure and the subsequent steps of recording the sequence variances and analyzing their frequency and structural and functional consequences.

**Example 37: Variance detection by sequencing.**

Sequencing by the Sanger dideoxy method or the Maxim Gilbert chemical cleavage method is widely used to determine the nucleotide sequence of genes. Presently, a worldwide effort is being put forward to sequence the entire human genome. The Human Genome Project as it is called has already resulted in the identification and sequencing of many new human genes. Sequencing can not only be used to identify new genes, but can also be used to identify variations between individuals in the sequence of those genes.

The following are the major steps involved in identifying sequence variations in a candidate gene by sequencing:

295

232/116

1. Amplification by the polymerase chain reaction (PCR) of 400-700 bp regions of the candidate gene from a panel of DNA samples. The DNA samples can either be cDNA or genomic DNA and will represent some cross section of the world population.
2. Sequencing of the resulting PCR fragments using the Sanger dideoxy method. Sequencing reactions are performed using fluorescently labeled dideoxy terminators and electrophoresed on an ABI 377 sequencer or its equivalent.
3. Analysis of the resulting data from the ABI 377 sequencer using software programs designed to identify sequence variations between the different samples analyzed.

A more detailed description of the procedure is as follows:

A candidate gene sequence is downloaded from an appropriate database. Primers for PCR amplification are designed which will result in the target sequence being divided into amplification products of between 400 and 700 bp. There will be a minimum of a 50 bp of overlap not including the primer sequences between the 5' and 3' ends of adjacent fragments to ensure the detection of variances which are located close to one of the primers.

Optimal PCR conditions for each of the primer pairs is determined experimentally. Parameters including but not limited to annealing temperature, pH,  $MgCl_2$  concentration, and KCl concentration will be varied until conditions for optimal PCR amplification are established. The PCR conditions derived for each primer pair is then used to amplify a panel of DNA samples (cDNA or genomic DNA) which is chosen to best represent the various ethnic backgrounds of the world population or some designated subset of that population.

PCR reactions are purified using the QIAquick 8 PCR purification kit (Qiagen cat#

28142) to remove nucleotides, proteins and buffers. The PCR reactions are mixed with 5 volumes of Buffer PB and applied to the wells of the QIAquick strips. The liquid is pulled through the strips by applying a vacuum. The wells are then washed two times with 1 ml of buffer PE and allowed to dry for 5 minutes under vacuum.

5 The PCR products are eluted from the strips using 60 ul of elution buffer.

The purified PCR fragments are sequenced in both directions using the Perkin Elmer ABI Prism™ Big Dye™ terminator Cycle Sequencing Ready Reaction Kit (Cat# 4303150). The following sequencing reaction is set up: 8.0 ul Terminator Ready  
10 Reaction Mix, 6.0 ul of purified PCR fragment, 20 picomoles of primer, deionized water to 20 ul. The reactions are run through the following cycles 25 times: 96°C for 10 second, annealing temperature for that particular PCR product for 5 seconds, 60°C for 4 minutes.

15 The above sequencing reactions are ethanol precipitated directly in the PCR plate, washed with 70% ethanol, and brought up in a volume of 6 ul of formamide dye. The reactions are heated to 90°C for 2 minutes and then quickly cooled to 4°C. 1 ul of each sequencing reaction is then loaded and run on an ABI 377 sequencer.

20 The output for the ABI sequencer appears as a series of peaks where each of the different nucleotides, A, C, G, and T appear as a different color. The nucleotide at each position in the sequence is determined by the most prominent peak at each location. Comparison of each of the sequencing outputs for each sample can be examined using software programs to determine the presence of a variance in the  
25 sequence. One example of heterozygote detection using sequencing with dye labeled terminators is described in Pui-Yan Kwok *et. al.* (Pui-Yan Kwok, Christopher Carlson, Thomas D. Yager, Wendy Ankener, and Deborah A. Nickerson, *Genomics* 23, 138-144 (1994)). The software compares each of the normalized peaks between all the samples base by base and looks for a 40% decrease in peak height and the concomitant

appearance of a new peak underneath. Possible variances flagged by the software are further analyzed visually to confirm their validity

5

**Example 38. Loss of heterozygosity.**

Loss of chromosomes or segments of chromosomes in disease cells results in loss of alleles in the disease cells compared to normal diploid cells. Such allele losses are a common occurrence in cancer, where they have been documented in over 1,500 publications in the past 14 years. More recent work has documented the occurrence of allele loss in other proliferative diseases. Several cytogenetic and molecular techniques have been developed to measure chromosome losses. The molecular techniques are preferable for identification of allele loss because they also show which allele is lost, and are therefore best suited to provide the information needed to implement the present invention.

In order to measure chromosome loss using molecular techniques it is necessary to be able to distinguish the paternally and maternally inherited copies of a given chromosome. DNA variances allow the two copies of a given chromosome to be distinguished because different alleles can be resolved electrophoretically. The standard method for analyzing allele loss in cancer is to compare tumor cell DNA with normal cell DNA, either in a Southern blot or using PCR based techniques. A patient's tumor DNA is said to be "informative" for allele loss only at loci where the patient's normal cells are heterozygous. When such heterozygous loci are examined in tumor cells often only one allele is detected. Such tumor cells have lost the heterozygous state which characterizes all normal somatic cells of the patient, hence the term loss of heterozygosity (LOH).

Several effective molecular procedures have been developed to measure LOH. These procedures have been applied most extensively to cancer tissues, however the same methods are effective in the study of nonmalignant diseases such as atherosclerotic plaques and endometriosis. The main steps are:

5

*1. Identify DNA variances at or near the locus to be investigated for LOH.*

10

LOH usually affects large segments of DNA, ranging from several megabases to an entire chromosome. As a result, accurate estimation of LOH at a specific locus can be obtained by measuring the frequency of LOH at neighboring polymorphic markers on the same chromosome, or more preferably on the same chromosome arm, or most preferably within several 10-20 megabases of the locus. However, to precisely measure LOH at a specific locus requires a variance at the locus. Different types of variances have been used to study LOH, including single nucleotide polymorphisms (SNPs), specifically SNPs that alter restriction endonuclease cleavage sites, called RFLPs. (For details of this approach see Vogelstein, B., et al., Allelotype of colorectal carcinomas. *Science* 244: 207-211, 1989). Also short tandem repeat polymorphisms (STRPs), including di-, tri- and tetranucleotide repeat polymorphisms have been used to measure LOH. (For details of this procedure see Jones and Nakamura, Deletion Mapping of Chromosome 3p in Female Genital Tract Malignancies Using Microsatellite Polymorphisms. *Oncogene* 7: 1631-1634, 1992.) Procedures for identifying variances are described in Examples 28, 29, 30 and 36.

15

20

25

*2. Prepare DNA from paired normal and disease tissue samples from patients being studied.*

Before preparing genomic DNA from tumor tissue it is important to assess tumor cell purity and viability, using microscopic examination of frozen sections if necessary. If embedded pathological specimens are being analyzed tumor cell purity can be

assessed by examining histologic sections before selecting areas for cell isolation and DNA purification. (See Johnson, et al., Direct Molecular Analysis of Archival Tumor Tissue for Loss of Heterozygosity, BioTechniques 19:190-191, 1995, and references therein for description of techniques for purifying tumor cell DNA from archival pathology samples.) Areas of necrosis and extensive admixture of normal and tumor tissue should be avoided. For Southern blotting ~5-10 ug of genomic DNA is required for each sample being analyzed. For PCR based methods as little as 5 to 10 ng of genomic DNA is sufficient; much less will suffice if two successive rounds of PCR amplification are used.

3. *Determine genotype in the normal and disease tissues using a quantitative or semi-quantitative procedure that allows the amount of each allele to be measured. Compare the ratio of alleles in the normal tissue to the ratio in the tumor tissue*

In order to show LOH at a given locus it is necessary to establish that the patient is constitutionally heterozygous at the locus. Thus DNA from normal tissue must be tested, either before or in parallel with tumor tissue DNA. A variety of methods can be used for quantitation of signal from the two alleles. If the alleles are compared on a Southern blot then signal in the bands corresponding to the two alleles can be counted by radioactive or nonradioactive techniques (see Ausubel, et al., Current Protocols in Molecular Biology, John Wiley & Sons). One method employs phosphor technology using a Molecular Dynamics PhosphorImager with ImageQuant software to measure signals. If the alleles are compared after PCR amplification then DNA sequencing can provide accurate quantitation of allele ratios. See, for example, Goldsborough and Kornberg, Allele-Specific Quantification of Drosophila Engrailed and Invected Transcripts, Proc. Natl. Acad. Sci. U.S.A. 91:12696-12700, 1994.

Using highly variable markers distributed across the genome a comprehensive map of LOH can be assembled for a specific cancer type. Such data sets have been termed allelotypes. Separate studies are necessary for different cancer (or other disease) types